



# The cost of mass gatherings during a pandemic

Ludovic Phalippou<sup>a</sup>, Dayin Zhang<sup>b,\*</sup>

<sup>a</sup> Said Business School, University of Oxford, Park End Street, Oxford, OX1 1HP, UK

<sup>b</sup> School of Business, University of Wisconsin-Madison, 975 University Avenue, Madison, WI, 53706, USA

## ARTICLE INFO

JEL classification:  
I12

Keywords:  
BLM protests  
COVID-19  
Rainfall  
Mass gathering

## ABSTRACT

In June 2020, the U.S., unlike other regions of the world, faced a surge in cases of COVID-19. Immediately prior to this wave of cases, the largest mass protests in U.S. history took place. We show that when holding other factors constant, COVID-19 cases increased most in places where more demonstrations occurred. We exploit variation in rainfall during the protest period as an exogenous source of variation in attendance. We find that good weather coincides with both more people protesting and more subsequent COVID-19 cases and deaths. Mass gatherings during a pandemic thus lead to more contraction and fatalities of COVID-19, and we quantify these effects.

## 1. Introduction

George Floyd was murdered during an arrest after a store clerk suspected he used a counterfeit \$20 bill in Minneapolis. Over the three weeks following this murder, an unprecedented wave of 5899 protests occurred across the U.S. At the peak of these protests, on June 6, about half a million people protested in 691 locations. A previous study (Madestam et al., 2013) shows that protests cause political change and can therefore lead to societal improvements. Anecdotally, following these protests, the issue of diversity and inclusion did become center stage in many corners of society, including at universities, creating hope for improvement.

These protests were notable due to their numbers and size and having occurred during a pandemic. As a result, questions have arisen about how these mass gatherings affect virus transmission. These concerns were voiced publicly and repeatedly. Dr. Anthony Fauci, director of the National Institute of Allergy and Infectious Diseases, said that “I get very concerned, as do my colleagues in public health when they see these kinds of crowds.” Moreover, Dr. Fauci suggested, “[i]t’s a perfect setup for further spread of the virus in the sense of creating these blips which might turn into some surges” (WTOP News, 2020). This paper is to test this hypothesis.

The previous literature finds mixed results for the relationship between mass gatherings and COVID-19 spread. Like us, Neyman and Dalsey (2021) and Dave et al. (2023) investigate the Black Lives Matter (BLM) protests but they find either small or no correlation between the protests and COVID-19 spread. This is in contrast to the empirical

findings in other mass gathering events, including Donald Trump rallies (Bernheim et al., 2020), anti-lockdown protests in Germany (Lange & Monscheuer, 2021), festivals and sporting events in Spain (Domènech-Montoliu et al., 2021; González-Val & Marcén, 2022), and religious gatherings in Malaysia (Che, Fazila, Edinur, Mohammad Khairul Azhar Abdul, & Safuan, 2020). Our paper explores exogenous variations in weather conditions to overcome a potential omitted variable bias, and finds a significant relationship between mass gatherings and virus transmission.

A descriptive assessment of aggregate COVID-19 cases in Fig. 1 shows a striking pattern. During the first wave of January to May 2020, the number of COVID-19 cases in the U.S. was nearly identical to Europe, a similar economy in terms of both development and population size. On May 31, 2020, the total number of COVID-19 cases in the U.S. was 1,779,731, and the total number of COVID-19 cases in Europe was 1,959,810. In contrast, over the two months that followed this historical mass protest in the U.S. (June–July 2020), nearly three million cases were registered in the U.S. versus less than one million in Europe.

Our empirical study investigates the impact of BLM protests on the growth in COVID-19 cases across counties. The main challenge in capturing the causal relationship is that some unobservable social characteristics and political preferences may determine the number of protesters and the risk of catching the virus. For example, counties with a more digitally literate population may be more exposed to the national sentiments of the BLM movement, and at the same time are better at practicing social distancing through remote working and online shopping.

\* Corresponding author.

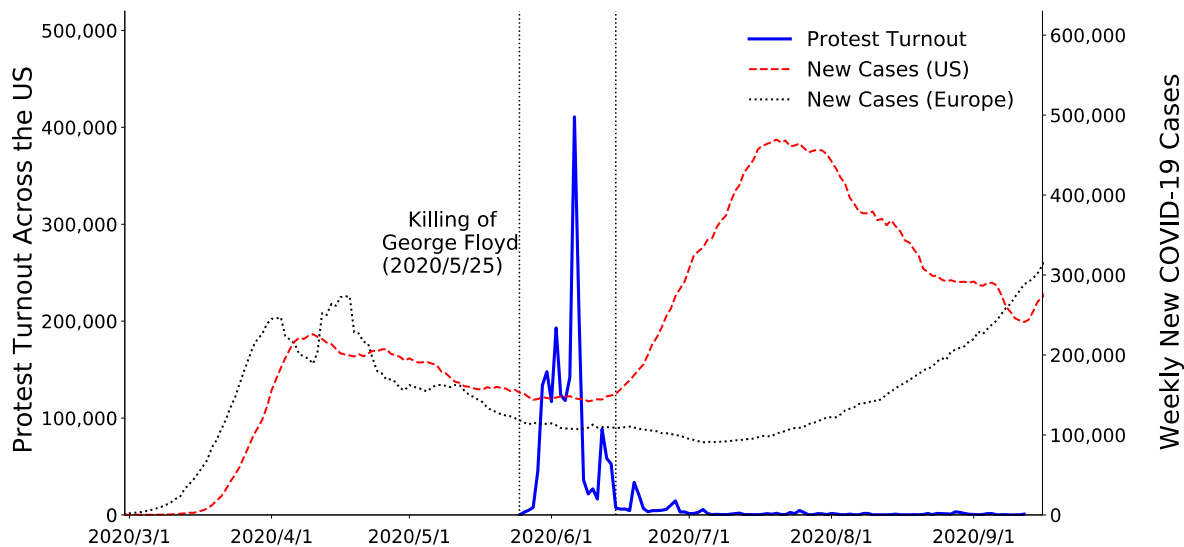
E-mail address: [dayin.zhang@wisc.edu](mailto:dayin.zhang@wisc.edu) (D. Zhang).

<https://doi.org/10.1016/j.ssmph.2023.101460>

Received 4 April 2023; Received in revised form 5 June 2023; Accepted 22 June 2023

Available online 26 June 2023

2352-8273/© 2023 The Authors. Published by Elsevier Ltd. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).



**Fig. 1.** BLM Protests and the COVID-19 Waves. *Note:* This figure plots the time series of BLM protest turnout (blue solid), and weekly new COVID-19 cases in both the US (red dash) and Europe (black dotted). The BLM protest turnout averages the reported participants in the BLM protests from three data sources: The Armed Conflict Location & Event Data Project (ACLED), Crowd Counting Consortium (CCC), and Count Love (CL). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

To address this issue of unobservable variables, we employ an instrumental variable (IV) strategy that plausibly isolates exogenous variation in the number of protesters. As shown in the literature (Madestam et al., 2013; Yanagizawa-Drott & Madestam, 2011), rainfall is a good instrumental variable for protests. The reasoning is straightforward; people are more likely to participate in outdoor protests if it does not rain. And rainfall is a random event, arguably uncorrelated with other factors that affect the spread of the virus. Thus, following the literature, we choose variation in rainfall during the three weeks of the protests (5/26–6/15, 2020) as an instrumental variable.

In addition, we control for a battery of socio-economic variables proposed by Spiegel and Tookes (2021), including lagged COVID-19 case growth, county demographics, Internet coverage, other outdoor activities, as well as weather conditions. We also include state fixed effects and therefore account for differences in i) test availability and ii) restriction policies, which are the same at the state level. For the same reason, we cluster standard errors at the state level, which allows for arbitrary within-state correlation.

First, the results show evidence that more people have protested in counties where less rainfall occurred: a 10% increase in rainfall during the BLM protest period reduces protest turnout by 1.2%; this negative correlation is statistically significant, and robust across alternative measures of protest turnout. We also show that pre-protest COVID-19 case levels and growth are uncorrelated with the rainfall during the protests. Most demographic characteristics are also balanced across counties with different rainfall levels. These results indicate that rainfall is a valid instrumental variable.

Next, we show that case growth over the six weeks following the protests is related to the amount of rainfall during the protest. A 10% increase in rainfall during the BLM protest period causes a 1.2% reduction or 7.2 fewer protesters, consequentially reducing the number of new cases by 2.0 for every 100,000 people over the following six weeks after the protests. Combining the results from these two stages, the IV regression estimate shows that for 100 additional protesters, we count 28.6 more cases in the six weeks after the BLM protests.

The estimated protest effect with the IV approach is much more considerable than that estimated using the Ordinary Least Squares (OLS) approach, which indicates that for every 100 additional protesters, there are 3.1 additional COVID-19 cases in the six weeks after the protests. This suggests a substantial downward bias due to the omitted social characteristic and political preference controls. And this explains the

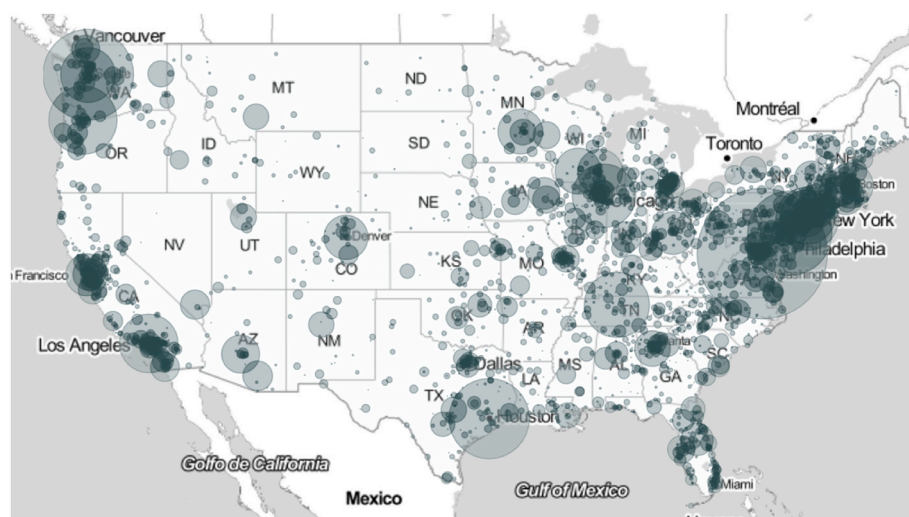
small or no correlation between the BLM protests and COVID-19 spread found by the previous literature (Dave et al., 2023; Neyman & Dalsey, 2021).

One threat to our IV approach is the potential violation of the exclusion condition. If the rainfall affects the COVID-19 spread through channels other than BLM protests, the causal interpretation of the regression coefficient is not valid. We test other channels by regressing the subsequent COVID-19 spread on the rainfall at non-protest times. These placebo tests show insignificant results, which indicate that rainfall does not strongly affect COVID-19 case growth through channels other than mass gatherings during BLM protests. This is not surprising, since most of the social events and gatherings are prohibited by shelter-in-place orders or voluntarily avoided during this period.

Finally, we show that massive gatherings of protesters during a pandemic cost human lives. In terms of fatalities, we estimate that 100 more protesters coincide with 0.4 more deaths.

This paper contributes to the literature that studies the risk factors for infectious diseases against the backdrop of the COVID-19 pandemic. Mass gatherings have been documented to play a crucial role in determining the basic reproduction rate  $R_0$  of the disease, which reflects the efficiency of its transmission (Domènech-Montoliu et al., 2021; Ebrahim & Memish, 2020). In spite of its importance in public health policy decisions, the estimation of  $R_0$  of COVID-19 had ranged widely (Sy et al., 2021; Liu et al., 2020). The idiosyncrasies in mass gathering events are argued to be one contributing factor in the wide range of the estimation (CDC COVID-19 Response Team et al., 2020). The Black Lives Matter (BLM) protests after the killing of George Floyd are the largest mass gathering events during the COVID-19 pandemic, so serve as the best laboratory to study this risk factor. The previous studies (Dave et al., 2023; Neyman & Dalsey, 2021) perform correlation analysis to examine the relationship between BLM protests and COVID-19 case growth and find at most a small correlation. This paper, however, exploits exogenous variations in weather conditions to make a causal inference for this critical question and achieve a contrast conclusion.

Our study is also related to the fast-growing literature on the economic impacts of COVID-19. Brodeur et al. (2021) counts more than 200 working articles from *National Bureau of Economic Research* (NBER) focusing on this issue in 2020 only, and surveys the related papers. As to the empirical studies in this line of literature (Chang & Meyerhoefer, 2021; Chetty et al., 2020; Kapteyn et al., 2020), the majority of them employ simple correlation analysis to provide timely economic analysis



**Fig. 2.** BLM protests across the U.S. *Note:* This figure plots the 5899 BLM protests across the United States in the 3-week period following the murder of George Floyd. Each green dot represents a county with protests, and the radius of the dots is proportional to the cumulative protest turnout during this 3-week period. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)

and insights to the public. For the follow-up causal analysis, this paper can provide exogenous variations of COVID-19 case growth across U.S. counties in the summer of 2020.

The rest of this paper proceeds as follows. Section 2 describes the background of the Black Lives Matter movement that we employ as a laboratory for studying how mass gathering affects COVID-19 growth. Section 3 describes our empirical approach. Section 4 presents the empirical results. Section 5 concludes.

## 2. Background and data

### 2.1. Murder of George Floyd and Black Lives Matter protests

On May 25, 2020, George Floyd, a 46-year-old black man, was murdered by a white police officer, Derek Chauvin, in Minneapolis while being arrested for using a counterfeit \$20 bill. In the following weeks, this murder sparked renewed protests in support of the Black Lives Matter (BLM) movement. The scale of these protests is unprecedented in U.S. history.

We obtain BLM protest information from three different sources: The Armed Conflict Location & Event Data Project (ACLED), Crowd Counting Consortium (CCC), and Count Love (CL). All three sources collect dates, locations, claims, and estimated attendance of media-reported protest events in the U.S. through the web crawling of local media. Each day, we aggregate the protest turnout by county from each data source and take the mean across all three. This should reduce measurement error in the data collection process.

In our empirical analysis, we define the BLM protest period as the three-week time range from May 26 (the day after George Floyd's death) to June 15, after which BLM protests turned from a national movement to sporadic local events with much lower attendance and media coverage, as shown in Fig. 1. In July, the total protest turnout was only 5% during the defined BLM movement period. This ratio dropped to 3% in August. We measure gatherings for BLM protests in each county by its cumulative turnout per 100,000 population during the BLM movement period. This is our key independent variable.

In Fig. 2, we map out the massive gathering during BLM protests at each protest location. We note that protests are widely spread across the nation. In the 3102 counties in the contiguous United States (i.e., excluding Alaska and Hawaii), nearly 50% of counties witnessed at least one BLM protest. The total turnout across the country is estimated to be 1.95 million. The upper panel of Table 1 shows that each U.S. county

has, on average, 8810 protesters attending 16 mass gatherings on the street. That translates to 605 protest turnouts per 100,000 population during this three-week protest period.

### 2.2. The COVID-19 pandemic waves in the U.S. And case data

The first American COVID-19 case was reported on January 20, 2020. Soon after, the virus rapidly spread to all corners of the nation. Cases surpassed 1 million on April 28, which marked the peak of the first wave of COVID-19 in the U.S., as shown in Fig. 1. This infection pattern matches one of a similar economy: Europe.

However, a second wave, much larger than the first wave, hit the U.S. but not Europe in June 2020, with more than 50,000 daily new cases for most of July and August. During the same period, Europe had very few cases.

We collect the county-level daily (cumulative) COVID-19 cases and deaths from Johns Hopkins University and summarize the data in Table 1. Before the BLM movement, a typical U.S. county had 183 new cases per 100,000 population in the three weeks from April 14 to May 4, and.

147 new cases from May 5 to May 25. Right before the outbreak of the BLM movement on May 25, 2020, a typical U.S. county had 514 cumulative COVID-19 cases per 100,000 population in the first 3 months of the pandemic. This number more than doubled after millions of protesters hit the street, with 254 new cases per 100,000 population registered from June 16 to July 6, and another 410 new cases from July 7 to July 27.

### 2.3. Rainfall and other weather data

We use rainfall information collected by the PRISM Climate Group, which provides daily precipitation estimates at any location in the U.S. using weather station observations. We aggregate the precipitation estimates to the county level by taking the spatial average.

Table 1 presents summary statistics for the rainfall data of the observed counties. On average, the daily rainfall in the U.S. counties during the BLM protest period is 0.11 inches. We notice a wide range of variations in rainfall levels across the nation. At least 5% of the population does not have any rain in the three-week protest period, and 5% of the population on the other extreme experiences almost 4 times the level of rainfall compared with the national average. We see these rainfall variations as random events. And we leverage these rainfall variations to

**Table 1**  
Summary statistics of variables.

	Mean	S.D.	P5	P25	P50	P75	P95	Obs.
<b>During BLM Protests</b>								
Number of BLM Protests	16	24	0	2	8	19	59	3102
Protest Turnout	8800	18,546	1	260	1995	6266	66,093	3102
Protest Turnout per 00,000'	605	1011	3	148	357	656	1913	3102
Rainfall <sup>BLM</sup> (thousandth inch)	110	140	0	22	68	130	433	3102
Median outdoor minutes	71.64	37.64	20.47	44.72	61.63	94.67	146.70	3102
Reopening Phase (standard)	1.25	1.01	0	0	1	2	3	3102
<b>Pre-BLM Protests</b>								
Cases per 00,000' @5/25/2020	513	657	44	136	261	582	1929	3102
Case Growth per 00,000':								
$t = -2$ (4/14–5/4, 2020)	183	271	9	36	84	205	801	3102
$t = -1$ (5/5–5/25, 2020)	147	163	8	42	88	196	446	3102
<b>Post-BLM Protests</b>								
Case Growth per 00,000':								
$t = 1$ (6/16–7/6, 2020)	254	244	28	79	154	384	752	3102
$t = 2$ (7/7–7/27, 2020)	410	367	56	138	300	590	1085	3102
Death Growth per 00,000':								
$t = 1$ (6/16–7/6, 2020)	4	6	0	1	3	6	12	3102
$t = 2$ (7/7–7/27, 2020)	6	7	0	2	4	8	18	3102
<b>Weather Controls:</b>								
Ave temperature (°F)	76.09	6.49	62.48	73.19	76.31	79.90	84.79	3102
Hot and humid weekday (%)	4.99	19.16	0.00	0.00	0.00	0.00	46.67	3102
Hot and humid weekend (%)	5.41	19.57	0.00	0.00	0.00	0.00	50.00	3102
Cold weekday (%)	0.55	3.37	0.00	0.00	0.00	0.00	3.33	3102
Cold weekend (%)	0.29	2.85	0.00	0.00	0.00	0.00	0.00	3102
<b>Demographics and Others</b>								
Black (%)	14.11	12.96	1.30	4.60	9.70	20.00	41.30	3102
Non-White Hispanic (%)	6.23	6.44	0.50	1.80	3.70	8.60	20.50	3102
Asian (%)	6.37	6.70	0.70	1.90	4.30	8.00	20.20	3102
Native American (%)	1.69	3.40	0.50	0.80	1.00	1.60	3.70	3102
Other non-white ethnicity (%)	5.84	6.06	0.50	1.70	3.50	8.00	19.50	3102
Per capita income (\$000')	34.09	9.00	22.37	28.22	32.76	38.07	52.23	3102
Population density (000' per mi <sup>2</sup> )	2.16	6.78	0.03	0.16	0.55	1.64	5.43	3102
Housing density (000' per mi <sup>2</sup> )	0.92	3.26	0.01	0.07	0.23	0.67	2.29	3102
Internet coverage (%)	82.74	6.68	70.20	79.70	83.90	87.20	91.50	3102
Age 65+ (%)	15.64	3.96	10.50	13.20	15.00	17.40	23.10	3102
Democrat voting	0.54	0.50	0.00	0.00	1.00	1.00	1.00	3102
Days since 1st Case	97	18	78	89	95	100	142	3102

*Note:* The unit of analysis is a county. *Rainfall<sup>BLM</sup>* is measured as daily weighted average precipitation in thousandth inch for the counties during the BLM period. Each day is weighted by the national aggregate BLM protest turnout to reflect the BLM protest intensity. *Median outdoor minutes* count median minutes people spend away from home measured by *SafeGraph*, and are averaged with national aggregate BLM protest turnout as weights. *Reopening Phase (standard)* indicates the reopening phase during BLM protests. COVID-19 case growth is defined as new cases per 100,000 population during specified period, and COVID-19 case is defined as the cumulative cases per 100,000 population at specified point of time. *Hot and humid weekday* and *Hot and humid weekend* are the percentage of weekdays and weekend days (respectively) in which the average temperature is above 80°F and the dew point is above 60°F. *Cold weekday* and *Cold weekend* are percentage of weekdays and weekend days in which the average high temperature is below 60°F. *Black*, *Non-White Hispanic*, *Asian*, *Native American*, and *Other*, are the fraction of the county's population that are Black, Non-White Hispanic, Asian, Native American, and other races (respectively). *Per capita income* is the per capita income in the county. *Population density* is defined as total population divided by the land square miles of the county. *Housing Density* is defined as the total number of homes in the county divided by residential land area. *Internet coverage* is the fraction of the population with access to the Internet. *Age 65+* is the fraction of the population that are over the age of 65. *Democrat voting* is based on the presidential election in 2016, and equal to 1 if Hillary Clinton won more votes than Donald Trump in 2016. *Days since 1st case* is the number of days since the first reported case of COVID-19 in a county.

capture the exogenous variations of protest turnouts and make causal inferences on the COVID-19 spread.

In addition, we collect other weather data to control for the contemporaneous weather conditions in the post-protest period. Some epidemiological studies believe these weather conditions directly or indirectly affect the COVID-19 spread (Karimi et al., 2020; McClymont & Hu, 2021). These conditions are either temperature or humidity related, and specified in the form of average temperature, hot and humid weekday percentage, hot and humid weekend percentage, cold weekday percentage, and cold weekend percentage.

#### 2.4. Outdoor activity measures

One concern with any IV estimation is the potential violation of the exclusion condition. Rainy weather reduces protester turnout and makes people less willing to attend other outdoor activities, such as BBQ parties. While there is no perfect measure of social outdoor activities, we resort to a proxy from *SafeGraph*. *SafeGraph* measures the time people

spend at or away from home using mobile phone GPS data, and reports the median minutes of the day people spend away from home at the census block group level. However, the data does not distinguish between social and individual activities and therefore only represent a proxy for the degree to which people stayed at home.

#### 2.5. Local reopening policy measures

We also control for the local reopening policy measures. It is important in our setting because the local policy is likely to be correlated with how rainfall interacts with protest turnouts. On the one hand, counties with slower reopening paces (stricter lockdowns) may have more people going to protests out of boredom rather than any real devotion to the political advocacy. If it rains, these less motivated protesters may be more likely to stay home. On the other hand, stricter lockdowns may make protesters more tolerant of staying out in the rain to protest since their alternative indoor activities are limited.

In our regression specifications, we add the state fixed effects to



control for the state-level COVID-19 containing and economic reopening policies. But there remain within-state policy variations.<sup>1</sup> They arise for two reasons. First, while many reopening orders are announced and implemented at the state level, there also are some adjustments to the effective dates across different counties. Furthermore, some local governments, especially in large metropolitan areas, have the autonomy to control their own pace of reopening.

To account for the within-state variations in COVID-19 containing and economic reopening policies, we gather data on the reopening status of different counties, including their reopening plans and effective dates, from various sources including the New York Times<sup>2</sup> and local government websites. We take a narrative approach and standardize the reopening plans from different local governments into a 5-phase procedure.

- Phase 0 (lockdown): Only essential businesses (e.g., healthcare) allowed;
- Phase 1: Outdoor businesses (e.g., construction) and manufacturing open, delivery or curbside pickup for retails and restaurants;
- Phase 2: Office-based businesses open, outdoor dining at restaurants;
- Phase 3: Personal care services (e.g., nail salons) open, indoor dining with capacity up to 50%;
- Phase 4: Places with large crowds (e.g., gyms, movie theaters) open;
- Phase 5: All businesses open.

## 2.6. Other demographic characteristics

We also control for demographic characteristics using data from the 2015–2019 five-year American Community Survey (ACS), and political preference using the 2016 presidential election data from MIT Election Data and Science Lab. The lower panel of Table 1 summarizes these county characteristics.

## 3. Research design

We aim to estimate the following specification relating COVID-19 case growth to the BLM protest turnout:

$$\text{Case Growth per } 100,000_c^{\text{Post BLM}} = \beta \log(\text{Protest Turnout per } 100,000_c^{\text{BLM}}) + X_c \gamma_0 + \mu_{\text{state}(c)} + \varepsilon_c, \quad (1)$$

where  $\text{Case Growth per } 100,000_c^{\text{Post BLM}}$  measures the new registered COVID-19 cases per 100,000 population in county  $c$  during the six-week period post BLM protests (6/16–7/27, 2020), and  $\text{Protest Turnout per } 100,000_c^{\text{BLM}}$  is the total turnout of BLM protesters per 100,000 population in county  $c$  during the three-week BLM protest period (5/26–6/15, 2020). We control for lagged COVID-19 growth, socio-economic demographics, weather conditions, and state fixed effects in  $X_c$  and  $\mu_{\text{state}(c)}$ . The set of explanatory variables follows the specification proposed by Spiegel and Tookes (2021). The regressions are weighted by county population, and the standard errors are clustered at the state level. Clustering at the state level allows for arbitrary within-state correlation and assumes no cross-state correlation.

We cannot use the OLS regression to recover the causal relationship

<sup>1</sup> During the BLM protests, most of the policy variations are at the state level, which is already controlled for by the state fixed effects. Specifically, 85% of the variations of the variable *Reopening Phase (standard)* are at the state level.

<sup>2</sup> We refer to the information on <https://www.nytimes.com/interactive/2020/us/states-reopen-map-coronavirus.html>. This page updates almost daily before its latest update on July 1, 2021. The historical snapshots are still available on some Internet archives, including WayBak Machine (<https://web.archive.org/>).

between the BLM protest turnout and COVID-19 case growth. Some unobservable social characteristics and political preferences may determine the number of protesters and the risk of catching the virus. For example, counties with a more digitally literate population may be more exposed to the national sentiments of the BLM movement, and at the same time are better in practicing social distancing through remote working and online shopping. The omitted variable bias could contaminate the causal interpretation of simple OLS estimation.

### 3.1. Instrumental variable (IV) regression

Our proposed solution to address the endogeneity concern above is to employ the exogenous variations in the protest turnout driven by the rainfall during the BLM protest period. Following the political economy literature (Madestam et al., 2013; Yanagizawa-Drott & Madestam, 2011), we exploit the fact that people are less likely to protest on the street if it rains, and use the rainfall level during the BLM protest period as the instrumental variable for the protest turnout. With the assumption that the level of rainfall during the protest period only affects the COVID-19 spread through protest attendance, we can causally estimate the effect of BLM protests on the COVID-19 case growth in the post-protest period.

**Construction of the Instrumental Variable** In our baseline specification, the instrumental variable is defined as the daily average rainfall during the BLM protest period. Fig. 1 shows the intensity of the BLM protests, which is highly related to people's emotions, is different during the three-week period following the killing of George Floyd. Hence, the weather condition on different days potentially has different effects on the number of protests. To capture the heterogeneity of nationwide protest intensity, we weight the rainfall levels by the total national BLM protest turnout when taking the average. The benefit of weighting is to improve the statistical power of the instrumental variable, but our results do not hinge on this weighting scheme. Specifically,

$$\text{Rainfall}_c^{\text{BLM}} = \frac{\sum_{d=5/26}^{6/15} w_d \text{Rainfall}_{c,d}/2020}{\sum_{d=5/26}^{6/15} w_d} \quad (2)$$

where the weight  $w_d$  is the total national BLM protest turnout on day  $d$  of 2020. The benefit of weighting is to improve the statistical power of the instrumental variable, even though our results do not hinge on this weighting scheme.

**Regression Specification** We first examine whether rainfall decreases the protest turnout by regressing the number of protesters on the level of rain during the same period. The first-stage regression is specified as

$$\log(\text{Protest Turnout per } 100,000_c^{\text{BLM}}) = \theta \log(\text{Rainfall}_c^{\text{BLM}}) + X_c \gamma_1 + \mu_{\text{state}(c)} + v_c. \quad (3)$$

where we control for the full set of variables listed in equation (1).

The coefficient of interest  $\theta$  measures the elasticity of the BLM protest turnout to the rainfall and is expected to be negative, since more rain leads to less protest turnout.

Here we control for a battery of socio-economic variables and the state fixed effects. These controls are not necessary for the purpose of identification if rainfall is unrelated to other determinants of the COVID-19 outcomes, but will help reduce the residual variation and improve the precision of our estimates.

In the second stage of our estimation, we investigate whether the rainfall affects the COVID-19 cases in the following six weeks, with the following reduced-form regression:

**Table 2**

OLS estimation: protest turnout &amp; post-protest case growth.

	Case Growth <sub>t = 1</sub> + Case Growth <sub>t = 2</sub> (per 00,000') (6/16–7/27, 2020)					
	(1)	(2)	(3)	(4)	(5)	(6)
log(Protest Turnout per 00,000')	29.95*** (5.02)	18.50*** (5.68)				
log(Protest Turnout)			28.98*** (8.17)	22.46*** (6.90)		
BLM Protest					3.59*** (4.77)	2.27*** (3.81)
Case per 00,000' @5/25		−0.07* (1.69)		−0.06 (−1.49)		−0.05 (−1.25)
Case Growth <sub>t = -1</sub>		0.33** (2.60)		0.32** (2.65)		0.31** (2.53)
Case Growth <sub>t = -2</sub>		−0.08 (−0.62)		−0.13 (−1.02)		−0.16 (−1.39)
Days since 1st Case		110.18*** (4.27)		87.77*** (3.35)		88.53*** (3.58)
Black		−0.04 (−0.03)		−0.33 (−0.23)		0.25 (0.17)
Non-white Hispanic		−2.44 (−0.12)		0.14 (0.01)		2.27 (0.13)
Asian		−10.49*** (−3.30)		−10.32*** (−3.32)		−10.43*** (−3.21)
Native American		−1.60 (−0.63)		−0.77 (−0.34)		−0.76 (−0.30)
Other non-white ethnicity		15.98 (0.85)		12.72 (0.67)		7.80 (0.45)
Median outdoor minutes		−0.94 (−1.49)		−0.45 (−0.73)		−0.97 (−1.50)
Reopening Phase (standard)		−6.11 (−0.22)		−5.07 (−0.19)		3.71 (0.12)
Population density		−12.21* (−1.91)		−12.75** (−2.12)		−11.08* (−1.70)
Housing density		16.88 (1.26)		17.24 (1.35)		13.25 (1.01)
Age 65+		−20.44*** (−6.81)		−18.84*** (−6.89)		−20.86*** (−6.49)
Democrat voting		20.67 (0.74)		8.51 (0.31)		24.78 (0.83)
Per capita income		2.40 (1.25)		2.21 (1.19)		2.46 (1.33)
Internet coverage		−8.50*** (−3.61)		−9.40*** (−3.79)		−7.26*** (−2.84)
Ave temperature		16.55*** (4.95)		16.39*** (4.92)		16.21*** (4.41)
HotHumid weekday		−1.37 (−0.45)		−1.75 (−0.60)		−1.62 (−0.58)
HotHumid weekend		−0.60 (−0.22)		−0.17 (−0.06)		0.39 (0.15)
Cold weekday		2.14 (0.46)		2.12 (0.44)		3.19 (0.65)
Cold weekend		9.73*** (3.33)		9.77*** (3.22)		9.45*** (3.09)
State FE	✓	✓	✓	✓	✓	✓
R <sup>2</sup>	0.73	0.82	0.74	0.82	0.74	0.82
Obs.	3101	3101	3101	3101	3101	3101

Note: This table shows the regressions in which we regress case growth per 100,000 population in the six-week window post BLM protests (6/16–7/27, 2020) on the three measures of BLM protest intensity: i) the log of protest turnout per 100,000 population (columns (1) and (2)), ii) the log of protest turnout (columns (3) and (4)), and iii) the number of the BLM protests (columns (5) and (6)). We winsorize both case growth per 100,000 population and BLM protest measures at the 5% level to limit the influence from extreme values. Each observation is a county. All regressions are weighted by county population, and control for state fixed effects. The standard errors are clustered at the state level. *t*-stats are shown in parentheses. \*, \*\*, \*\*\* indicates the observed coefficient is statistically significant at the 90%, 95%, and 99% confidence intervals, respectively.

$$\text{Case Growth per } 00,000_c^{\text{Post BLM}} = \kappa \log(\text{Rainfall}_c^{\text{BLM}}) + X_c \gamma_2 + \mu_{\text{state}(c)} + \varepsilon_c. \quad (4)$$

Combining the estimates in equations (3) and (4), we can produce the estimation of BLM protests' effect on the post-protest COVID-19 case growth with a two-stage least squares (2SLS) approach:

$$\text{Case Growth per } 00,000_c^{\text{Post BLM}} = \beta^{\text{IV}} \log(\text{Protest Turnout per } 00,000_c^{\text{BLM}}) + X_c \gamma_1 + \mu_{\text{state}(c)} + \sigma_c. \quad (5)$$

With the assumption that rainfall affects the later COVID-19 case growth only through the channel of BLM protest attendance (exclusion condition), we can avoid the endogeneity problem, and strictly interpret the coefficient  $\beta^{\text{IV}}$  as the causal effect of the BLM protests on the spread of COVID-19.

**Table 3**  
Balance test for county characteristics.

Dependent Variables	log(Rainfall <sup>BLM</sup> )		State FE	R <sup>2</sup>	Obs.
	Coeff.	t-stat.			
	(1)	(2)	(3)	(4)	(5)
<u>Pre-BLM COVID-19 Cases</u>					
Cases per 00,000' @5/25/2020	-3.08	(-0.13)	✓	0.54	3101
Case Growth per 00,000':					
$t = -1$ (5/5-5/25, 2020)	-5.85	(-0.85)	✓	0.40	3101
$t = -2$ (4/14-5/4, 2020)	-2.91	(-0.37)	✓	0.59	3101
<u>Demographics</u>					
Black (%)	-0.16	(-0.54)	✓	0.38	3101
Non-White Hispanic (%)	-0.86*	(-1.73)	✓	0.56	3101
Asian (%)	-0.07	(-0.49)	✓	0.46	3101
Native American (%)	0.24	(1.27)	✓	0.30	3101
Other non-white ethnicity (%)	-0.87*	(-1.85)	✓	0.59	3101
Per capita income (\$000')	0.17	(0.69)	✓	0.24	3101
Population density (000' per mi <sup>2</sup> )	0.17	(0.45)	✓	0.32	3101
Housing density (000' per mi <sup>2</sup> )	0.10	(0.53)	✓	0.27	3101
Internet coverage (%)	-0.05	(-0.15)	✓	0.24	3101
Age 65+ (%)	0.14*	(1.89)	✓	0.25	3101
Democrat voting	0.01	(0.29)	✓	0.23	3101
log(Days since 1st Case)	-0.01	(-1.43)	✓	0.21	3101

*Note:* This table shows the correlation between the log of rainfall with pre-protest COVID-19 case growth and county characteristics. We winsorize both case growth per 100,000 population and rainfall level (instrumental variable) at the 5% level to limit the influence from extreme values. Each observation is a county. All regressions are weighted by county population, and control for state fixed effects. The standard errors are clustered at the state level. *t*-stats are shown in parentheses. \*, \*\*, \*\*\* indicates the observed coefficient is statistically significant at the 90%, 95%, and 99% confidence intervals, respectively.

The threat to the identification is the violation of the exclusion condition. One possible violating channel is that the rainfall during the BLM protest period also negatively affects other outdoor activities (e.g., exercises, BBQs, etc.) and thus reduces the spread of COVID-19. To mitigate this concern, we first include the share of time people spend away from home during the BLM protests into our control variables to explicitly control for this channel in our baseline specification.

Another possible violation of the exclusion condition is through the persistence of the rainfall. If the counties that have more rain during the BLM protests also have more rain in the following three months, the reduced COVID-19 spread might be due to the contemporaneous weather effect. To address this concern, we also include the contemporaneous daily rainfall in the control variables.

#### 4. Empirical results

In this section, we present our empirical results on the causal relationship between the mass gathering in the BLM protests and the COVID-19 spread with the instrumental variable approach.

##### 4.1. Correlation between BLM protests and COVID-19 cases

We start by reporting the conditional correlation between the BLM protest turnout and the new COVID-19 cases in the post-protest period (6/16-7/27, 2020).

Columns (1), (3) and (5) of Table 2 shows results when we only control for state fixed effects. The difference between these three regressions is the proxy used to measure protest turnout. The effect is strongest when using an absolute measure: the total number of people who came to the protest. Columns (2), (4) and (6) show the results when we add all the control variables. The coefficient can be interpreted as a measure of the difference in case growth after the protests for two counties that would have similar i) pre-protest cases per capita, ii) demographics, and iii) climate, but different numbers of protesters.

Economic magnitudes are large: a county with 1% more protesters sees 0.185 more COVID-19 cases in the following six-week period. A

county, on average, has 605 protesters per 100,000 people, so a 1% increase corresponds to 6.05 more protesters. In other words, we can interpret the estimation in column (2), for example, as 100 additional protesters are associated with 3.1 ( $=0.185/6.05 \times 100$ ) more cases in the following six-week period.

In summary, the OLS regression shows a strong positive conditional correlation between protest turnout and post-period COVID-19 cases.

##### 4.2. Balance test of the instrumental variable

A fundamental identifying assumption for the IV approach is that rainfall during the BLM protests is random across counties, so it should be uncorrelated with other determinants of COVID-19 outcomes. We test this premise in Table 3 by regressing pre-protest COVID-19 case growth and demographic characteristics on the constructed rainfall instrumental variable. We report the correlation of rainfall with each dependent variable. We control for the state fixed effects and cluster the standard errors at the state level, as in the main regressions.

The upper panel shows the rainfall during BLM protests is not significantly correlated with any pre-protest COVID-19 outcomes. This is effectively the placebo test of the rainfall effect on pre-protest COVID-19 spread. In other words, the rainfall does not pick up areas regarding their COVID-19 situations.

The lower panel shows the correlation between the rainfall during BLM protests and the demographic characteristics. This is to test whether the heavy-rain and light-rain counties are balanced along these demographic dimensions. While most demographic characteristics are well balanced between heavy-rain and light-rain counties, non-white Hispanics, other ethnicities, and retired people slightly tend to live in counties with less rainfall during the BLM protest period. In the online appendix, we show this is probably driven by the historical migration pattern, that is driven by climate conditions. Once we control for the historical precipitation, the correlation is gone. The results in Table 3 lend credibility to our identification strategy by showing little correlation between the instrumental variable and all pre-determined county characteristics.

##### 4.3. The effect of rainfall on BLM protests

We first estimate the first stage regression with specification (3), and report the full estimation results in Table 4. Columns (1) and (2) estimate the effects of rainfall on our preferred independent variable, protest turnout per 100,000 population. Column (1) only controls for predetermined county characteristics, and column (2) also includes lagged COVID-19 case growth and contemporaneous weather conditions. The estimated rainfall effects on protest turnout are almost identical in both regression specifications. A 10% increase in the daily rainfall level would reduce BLM protest turnout by 1.2%, or 7.2 ( $=1.2\% \times 605$ ) protesters per 100,000 population. Rain strongly decreases attendance in the BLM protests and thus is an influential instrument variable for BLM protests. This effect of rainfall on BLM protests is significant both economically and statistically, with *F* statistics well above the rule-of-thumb non-weak IV threshold of ten (Stock et al., 2002). This finding implies that our results are not subject to weak-instrument bias.

The coefficients on the demographic controls show that counties with a higher percentage black population see more protesters. Democratic counties have larger protest turnouts as well. This is consistent with the political preference of these counties. We also find places with better access to the Internet to have more protesters since the BLM movement was initially sparked and further escalated on the Internet.

In columns (3) through (6), we estimate the rainfall's effect on two other measures of BLM protest: total protest turnout and the number of protests. The impact of rain is consistently negative and significant. This alleviates our concern about the misspecification of the protest measure.

**Table 4**

First stage regression: rainfall effect on BLM protest turnout.

	log(Protest Turnout per 00,000')		log(Protest Turnout)		Number of BLM Protests	
	(1)	(2)	(3)	(4)	(5)	(6)
log(Rainfall <sup>BLM</sup> )	-0.10*** (-3.27)	-0.12*** (-3.87)	-0.22*** (-4.26)	-0.23*** (-5.84)	-5.01*** (-3.35)	-5.47*** (-3.55)
Case per 00,000' @5/25		-0.00 (-0.08)		-0.00** (-2.38)		-0.01** (-2.21)
Case Growth <sub>t-1</sub>		-0.00 (-0.49)		0.00 (0.17)		0.01 (0.88)
Case Growth <sub>t-2</sub>		0.00 (1.01)		0.00*** (3.79)		0.04** (2.68)
Days since 1st Case		0.37*** (3.14)		1.27*** (3.16)		11.51* (1.68)
Black	0.03*** (4.07)	0.02*** (3.22)	0.06*** (4.16)	0.03*** (2.90)	0.20** (2.38)	0.02 (0.24)
Non-white Hispanic	-0.07 (-1.10)	-0.09 (-1.47)	-0.15* (-1.96)	-0.18*** (-2.81)	-2.42*** (-3.03)	-2.61*** (-3.25)
Asian	-0.00 (-0.46)	-0.02** (-2.55)	0.01 (0.64)	-0.03* (-1.97)	0.15 (1.35)	-0.32 (-1.44)
Native American	-0.00 (-0.29)	-0.01 (-1.31)	-0.02 (-0.91)	-0.04** (-2.47)	-0.10 (-1.21)	-0.35*** (-3.23)
Other non-white ethnicity	0.04 (0.62)	0.05 (0.68)	0.18** (2.16)	0.16** (2.09)	3.39*** (4.09)	3.34*** (4.18)
Median outdoor minutes		-0.01*** (-5.23)		-0.03*** (-9.80)		-0.12*** (-3.94)
Reopening Phase (standard)		-0.02 (-0.21)		-0.02 (-0.14)		-3.15* (-1.94)
Population density	-0.00 (-0.02)	-0.01 (-0.49)	0.06 (1.10)	0.02 (0.48)	0.50 (1.30)	-0.34 (-0.44)
Housing density	0.03 (0.45)	0.07 (1.42)	-0.06 (-0.73)	0.03 (0.38)	-0.15 (-0.18)	1.90 (1.16)
Age 65+	-0.04** (-2.14)	-0.04** (-2.60)	-0.11*** (-2.74)	-0.11*** (-3.86)	-0.29 (-1.63)	-0.23 (-1.41)
Democrat voting	0.78*** (4.97)	0.55*** (3.63)	1.59*** (5.34)	1.03*** (4.72)	5.13*** (3.51)	3.94** (2.26)
Per capita income	0.02** (2.08)	0.01 (0.88)	0.04*** (2.79)	0.01 (0.85)	0.37** (2.60)	0.00 (0.02)
Internet coverage	0.05*** (5.39)	0.04*** (4.70)	0.11*** (8.27)	0.07*** (4.63)	-0.30** (-2.58)	-0.21 (-1.07)
Ave temperature		-0.06*** (-3.76)		-0.04* (-1.95)		-0.45** (-2.33)
HotHumid weekday		0.01 (0.70)		0.01 (0.35)		-0.29 (-0.49)
HotHumid weekend		-0.00 (-0.31)		-0.01 (-0.35)		-0.05 (-0.13)
Cold weekday		0.00 (0.10)		0.00 (0.11)		-0.37 (-1.02)
Cold weekend		-0.00 (-0.03)		-0.00 (-0.05)		0.12 (0.35)
State FE	✓	✓	✓	✓	✓	✓
R <sup>2</sup>	0.41	0.44	0.64	0.68	0.59	0.69
Obs.	3101	3101	3101	3101	3101	3101

Note: This table shows the first stage regression analysis. The dependent variable is i) the log of protest turnout per 100,000 population (columns (1) and (2)), ii) the log of protest turnout (columns (3) and (4)), and iii) the number of the BLM protests (columns (5) and (6)). We winsorize the outcome variables and instrumental variable at the 5% level to limit the influence from extreme values. All regressions are weighted by county population, and control for state fixed effects. Standard errors are clustered at the state level. *t*-stats are shown in parentheses. \*, \*\*, \*\*\* indicates the observed coefficient is statistically significant at the 90%, 95%, and 99% confidence intervals, respectively.

#### 4.4. BLM protests lead to more COVID-19 infections

The main results of this paper are presented in Table 5. Column (1) implements the reduced form regression using the estimation equation (4). It shows that a 10% increase in the daily rainfall level during the BLM protest period reduces the number of new cases by 2.036 for every 100,000 people over six weeks after the protests. This is the consequence of a reduction of 7.2 protesters for every 100,000 people. Combining these two results, we can get the 2SLS estimate in column (2), implying that 100 more protesters will cause 28.63 (=1.7319/6.05\*100) more cases.

We also show the results if we measure the protest in alternative forms in columns (3) and (4). The significantly positive protest effect also holds.

#### 4.5. Comparison between OLS and IV estimates

Column (2) of Table 2 shows the OLS estimation of the effect of BLM protests on COVID-19 spread. The estimated effect is significantly positive, but quantitatively much smaller. For every 100 additional protesters, there are 3.1 additional COVID-19 cases in the next six-week period after the protests ceased. This finding is consistent with the findings in Neyman and Dalsey (2021) and Dave et al. (2023), where they find at most a small correlation between the BLM protests and COVID-19 spread.

The large discrepancy between the OLS and IV estimates indicates that the omitted variable bias is downward and substantial. The coefficients on the control variables in Tables 2 and 4 give some clue of the source of these omitted variables. For example, the Internet coverage



**Table 5**

IV estimate: protest effect on post-protest case growth.

	Case Growth <sub>t = 1</sub> + Case Growth <sub>t = 2</sub> (per 00,000') (6/16–7/27, 2020)			
	Reduced Form	2SLS	2SLS	2SLS
	(1)	(2)	(3)	(4)
log(Protest Turnout per 00,000')		173.19** (2.27)		
log(Protest Turnout)			87.35** (2.64)	
BLM Protest				3.72*** (3.98)
log(Rainfall <sup>BLM</sup> )	−20.36*** (−2.78)			
Control variables:				
Lagged Case Growth	✓	✓	✓	✓
Median outdoor minutes	✓	✓	✓	✓
Reopening Phase (standard)	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
Weather Condition	✓	✓	✓	✓
State FE	✓	✓	✓	✓
First Stage Diagnostics:				
Cragg-Donald Wald F		31.85	69.88	622.39
Kleibergen-Paap Wald Rk F		15.01	34.08	12.64
R <sup>2</sup>	0.82			
Obs.	3101	3101	3101	3101

Note: This table shows the main regression results. Column (1) is the reduced form regression, which shows the effect of rainfall on case growth. Columns (2)–(4) report the results of two-stage-least-squares (2SLS) regressions of case growth on the three different BLM protest measures. We winsorize i) case growth per 100,000 population, ii) BLM protest turnout per 100,000 population, and iii) rainfall level at 5% to limit the influence from extreme values. All regressions are weighted by county population, and control for state fixed effects. Standard errors are clustered at the state level. *t*-stats are shown in parentheses. \*, \*\*, \*\*\* indicates the observed coefficient is statistically significant at the 90%, 95%, and 99% confidence intervals, respectively.

has a strong positive correlation with BLM protest turnouts, but negative correlation with COVID-19 spread. We can see the Internet coverage as a noisy proxy for digital literacy. Counties with a more digitally literate population are more exposed to the national sentiments of the BLM movement, as well as local protest calls that are circulated in social media. Thus, these counties witness more.

Protest turnouts. At the same time, these counties have more people working remotely from work and shopping online, which help slow down the spread of COVID-19. Therefore, the omitted digital literacy contributes to the endogeneity bias of the OLS estimation.

The IV regressions help solve the omitted variable problem and uncover the correct causal effect of the BLM protest on the post-period cases.

#### 4.6. Placebo tests

To provide further evidence on the validity of the research design, we conduct a few placebo tests.

A valid instrumental variable should satisfy the exclusion condition. Otherwise, the causal interpretation of the IV regression coefficient will be contaminated. In our setting, one concern might be that the rainfall would affect the subsequent COVID-19 spread through channels other than BLM protests. For example, rainy weather might drive some events (birthday parties, weddings, etc.) from outdoor to indoor, which accelerates the transmission of COVID-19.

Our strategy to address this concern is to investigate the rainfall effect on COVID-19 spread (the reduced form regression) in periods other than during BLM protests. Panel A of Table 6 reports the results of two such placebo tests. In column (1), we shift the time window forward by 1 month, and regress the COVID-19 case growth during 7/16–8/27 on the

**Table 6**

Placebo test results Panel A: Placebo reduced form regression for rainfall effect on case growth (non-protest time windows).

	Case Growth (per 00,000') (7/16–8/27, 2020)	Case Growth (per 00,000') (8/16–9/27, 2020)
	Reduced Form	Reduced Form
	(1)	(2)
log(Rainfall): 6/26–7/15, 2020	3.96 (0.61)	
log(Rainfall): 7/26–8/15, 2020		8.38 (1.31)
Control variables:		
Lagged Case Growth	✓	✓
Median outdoor minutes	✓	✓
Reopening Phase (standard)	✓	✓
Demographics	✓	✓
Weather Condition	✓	✓
State FE	✓	✓
R <sup>2</sup>	0.78	0.62
Obs.	3101	3095

Panel B: Placebo test for the protest correlation with pre-protest case growth

	Case Growth (per 00,000')		
	<i>t</i> = −3 (3/24–4/13, 2020)	<i>t</i> = −2 (4/14–5/4, 2020)	<i>t</i> = −1 (5/5–5/25, 2020)
	(1)	(2)	(3)
log(Protest Turnout per 00,000') (5/26–6/15, 2020)	−1.25 (−0.59)	0.43 (0.25)	−1.15 (−0.95)
Control variables:			
Lagged Case Growth	✓	✓	✓
Median outdoor minutes	✓	✓	✓
Reopening Phase (standard)	✓	✓	✓
Demographics	✓	✓	✓
Weather Condition	✓	✓	✓
State FE	✓	✓	✓
R <sup>2</sup>	0.87	0.84	0.75
Obs.	3101	3101	3101

Note: Panel A regresses the subsequent case growth on the log of rainfall in July and August. In column (1), we shift the time window forward by 1 month, and regress the COVID-19 case growth during 7/16–8/27 on the rainfall during 6/26–7/15. In column (2), we shift the time window forward by 2 months, and regress the COVID-19 case growth during 8/16–9/27 on the rainfall during 7/26–8/15. Panel B regresses the case growth in the pre-protest periods on the BLM protest turnouts. Columns (1)–(3) look at the case growth during 3/24–4/13, 4/14–5/4, and 5/5–5/25 respectively. We winsorize i) case growth per 100,000 population, ii) BLM protest turnout per 100,000 population, and iii) rainfall level at 5% to limit the influence from extreme values. All regressions are weighted by county population, and control state fixed effects. Standard errors are clustered at the state level. *t*-stats are shown in parentheses. \*, \*\*, \*\*\* indicates the observed coefficient is statistically significant at the 90%, 95%, and 99% confidence intervals, respectively.

rainfall during 6/26–7/15. In column (2), we shift the time window forward by 2 months, and regress the COVID-19 case growth during 8/16–9/27 on the rainfall during 7/26–8/15. In both tests, rainfall does not predict COVID-19 case growth in the following six-week period, after controlling for the observable county characteristics. This result is consistent with the finding in the health literature (Karimi et al., 2020; McClymont & Hu, 2021), that precipitation is not correlated with COVID-19 transmission and mortality. Our interpretation is that most of the social events and gatherings are prohibited (concerts, restaurant dining, etc.) or voluntarily avoided (graduation commencements, etc.) during this period. Therefore, the rainy weather has a minimal effect in moving large crowds from outdoor to indoor, and thus does not show a significant direct effect on the COVID-19 spread.

We conduct another set of placebo tests in Panel B of Table 6, where we regress the case growth in the pre-protest periods on the BLM protest

**Table 7**  
Protest effect on post-protest death growth.

	Death Growth <sub>t = 1</sub> + Death Growth <sub>t = 2</sub> (per 00,000') (6/16–7/27, 2020)			
	Reduced Form (1)	2SLS (2)	2SLS (3)	2SLS (4)
log(Protest Turnout per 00,000')		2.41** (2.04)		
log(Protest Turnout)			1.21** (2.14)	
BLM Protest				0.05*** (2.97)
log(Rainfall <sup>BLM</sup> )	−0.28** (−2.10)			
Control variables:				
Lagged Case Growth	✓	✓	✓	✓
Median outdoor minutes	✓	✓	✓	✓
Reopening Phase (standard)	✓	✓	✓	✓
Demographics	✓	✓	✓	✓
Weather Condition	✓	✓	✓	✓
State FE	✓	✓	✓	✓
First Stage Diagnostics:				
Cragg-Donald Wald F		31.85	69.88	622.39
Kleibergen-Paap Wald Rk F		15.01	34.08	12.64
R <sup>2</sup> Obs.	0.65 3101	3101	3101	3101

Note: The table shows the Instrumental Variable (IV) regression results on COVID-19 death growth per 100,000 population in the six-week window post BLM protests (6/16–7/27, 2020). Column (1) is the reduced form regression, which shows the effect of rainfall on case growth. Columns (2)–(4) report the results of two-stage-least-squares (2SLS) regressions of case growth on the three different BLM protest measures. We winsorize i) death growth per 100,000 population, ii) BLM protest turnouts per 100,000 population, and iii) rainfall level at 5% to limit the influence from extreme values. All regressions are weighted by county population, and control for state fixed effects. Standard errors are clustered at the state level. *t*-stats are shown in parentheses. \*, \*\*, \*\*\* indicates the observed coefficient is statistically significant at the 90%, 95%, and 99% confidence intervals, respectively.

turnouts, controlling for observable county characteristics.

For the three periods leading up to the protest period ( $t = -3, -2, -1$ ), the correlation between protest turnout and case growth is insignificant. Because the BLM protests had not yet started, the protest turnout does not have any explanatory power on the cross-sectional distribution of the COVID-19 case growth. These results pass the placebo test in that the mass gathering is not correlated with the pre-protest period case growth.

#### 4.7. COVID-19 death toll of BLM protests

What is the cost of the protest-driven case growth in the second wave of the pandemic in the U.S.? This question can be partly answered by measuring the death toll due to the BLM protests. Table 7 presents the effect of BLM protests on new COVID-19 deaths in the six weeks following the outset of BLM protests. The reduced form regression in column (1) demonstrates that 1% more daily rainfall during the BLM protest period reduces COVID-19 deaths by 0.0028%. Column (2) indicates that 100 more protesters coincide with only 0.40 ( $=2.41/605 \times 100$ ) more deaths. And again, we see a similar pattern if you use alternative measures of death outcomes in columns (3) and (4).

The result in Table 7 measures the direct life cost of the mass gathering in the BLM protests. In addition, the economic toll of the extra spread of COVID-19 is also substantial, as shown by many studies (Albanesi & Kim, 2021; Alsan et al., 2021; Chetty et al., 2020).

## 5. Conclusion

In this paper, we have provided evidence that the mass gathering events during a pandemic helped the COVID-19 virus spread. This

finding proves that mass gathering is a significant risk factor that accelerates the spread of infectious respiratory diseases, including COVID-19. From a positive perspective, this helps explain the wide range of the estimated COVID-19 reproduction rate  $R_0$ . Despite the crucial importance of this parameter, early estimates range widely from 1.4 to 6.49 (Liu et al., 2020). This paper proposes a vital factor of human gathering behaviors, which should be considered to reduce the estimation noise of the basic reproduction rate. From a normative perspective, the direct linkage between mass gatherings and the accelerated spread of COVID-19 justifies the social distancing measures taken by governments around the world in the early stages, when effective vaccines had not been developed and the death rate was still high.

While our study focuses on the second wave of the COVID-19 surge in the U.S., this causal relationship generally applies to other time periods and countries. Anecdotal evidence in different settings tells a consistent story. At the very start of the pandemic, the Baibuting community in Wuhan, China, saw an accelerated spread of COVID-19 in late January 2020 after thousands of families gathered at the Lunar New Year banquet. Across the Pacific, New Orleans became one of the earliest hotspots for COVID-19 in the U.S., which public health officials have largely accepted is associated with its Mardi Gras celebration in February 2020.

## Ethics statement

I declare no conflict of interest. The writing of this commentary was not funded from any source.

## Author contributions

Each author contributed equally to this work.

## Data availability

Data will be made available on request.

## Acknowledgments

We thank Heejin Yoon and Qiulin Chen for their great research assistance.

## Appendix A. Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.ssmph.2023.101460>.

## References

- Albanesi, S., & Kim, J. (2021). Effects of the COVID-19 recession on the US labor market: Occupation, family, and gender. *The Journal of Economic Perspectives*, 35, 3–24.
- Alsan, M., Chandra, A., & Simon, K. (2021). The great unequalizer: Initial health effects of COVID-19 in the United States. *The Journal of Economic Perspectives*, 35, 25–46.
- Bernheim, B. D., Buchmann, N., Freitas-Groff, Z., & Otero, S. (2020). The effects of large group meetings on the spread of COVID-19: The case of Trump rallies. SSRN: <https://doi.org/10.2139/ssrn.3722299>.
- CDC COVID-19 Response Team, Bialek, S., Bowen, V., Chow, N., Curns, A., Ryan, G., Hall, A., et al. (2020). Geographic differences in COVID-19 cases, deaths, and incidence—United States. *Morbidity and Mortality Weekly Report*, 69, 465. February 12–April 7, 2020.
- Brodeur, A., Gray, D., Islam, A., & Bhuiyan, S. (2021). A literature review of the economics of covid-19. *Journal of Economic Surveys*, 35, 1007–1044.
- Chang, H.-H., & Meyerhoefer, C. D. (2021). Covid-19 and the demand for online food shopping services: Empirical evidence from taiwan. *American Journal of Agricultural Economics*, 103, 448–465.
- Che, M., Fazila, N., Edinur, H. A., Mohammad Khairul Azhar Abdul, R., & Safuan, S. (2020). A single mass gathering resulted in massive transmission of COVID-19 infections in Malaysia with further international spread. *Journal of Travel Medicine*, 27(3).
- Chetty, R., Friedman, J. N., Hendren, N., Michael Stepner, & The Opportunity Insights Team.. (2020). *The economic impacts of COVID-19: Evidence from a new public database built using private sector data*. national Bureau of economic research. Technical report.
- Domènech-Montoliu, S., Maria Rosario, P.-S., Vidal-Utrillas, P., Latorre-Poveda, M., Alba Del Rio-González, Ferrando-Rubert, S., Ferrer-Abad, G., Sánchez-Urbano, M.,

- Aparisi-Esteve, L., Badenes-Marques, G., et al. (2021). Mass gathering events and COVID-19 transmission in borriana (Spain): A retrospective cohort study. *PLoS One*, 16, Article e0256747.
- Ebrahim, S. H., & Memish, Z. A. (2020). COVID-19—the role of mass gatherings. *Travel Medicine and Infectious Disease*, 34, Article 101617.
- González-Val, R., & Marcén, M. (2022). Mass gathering events and the spread of infectious diseases: Evidence from the early growth phase of COVID-19. *Economics and Human Biology*, 46, Article 101140.
- Kapteyn, A., Angrisani, M., Bennett, D., Bruine de Bruin, W., Darling, J., Gutsche, T., Liu, Y., Meijer, E., Perez-Arce, F., Schaner, S., Thomas, K., & Bas, W. (2020). Tracking the effect of the covid-19 pandemic on the lives of american households. *Survey Research Methods*, 14, 179–186.
- Karimi, S. M., Majbourni, M., White, K., Little, B., Paul McKinney, W., & DuPre, N. (2020). Spring weather and COVID-19 deaths in the U.S., preprint, infectious diseases (except HIV/AIDS).
- Lange, M., & Monscheuer, O. (2021). Spreading the disease: Protest in times of pandemics. *Health Economics*, 31, 2664–2679.
- Liu, Y., Gayle, A. A., Wilder-Smith, A., & Rocklöv, J. (2020). The reproductive number of COVID-19 is higher compared to SARS coronavirus. *Journal of Travel Medicine*.
- Madestam, A., Shoag, D., Veuger, S., & Yanagizawa-Drott, D. (2013). Do political protests matter? Evidence from the tea party movement. *Quarterly Journal of Economics*, 128, 1633–1685.
- McClymont, H., & Hu, W. (2021). Weather variability and COVID-19 transmission: A review of recent research. *International Journal of Environmental Research and Public Health*, 18, 396.
- Neyman, G., & Dalsey, W. (2021). Black lives matter protests and COVID-19 cases: Relationship in two databases. *Journal of Public Health*, 43, 225–227.
- PRISM Climate Group. (2004). Oregon state university. <http://prism.oregonstate.edu>.
- Spiegel, M., & Tookes, H. (2021). Business restrictions and COVID-19 fatalities. *Review of Financial Studies*, 34, 5266–5308.
- Stock, J. H., Wright, J. H., & Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20, 518–529.
- Sy Karla Therese, L., White, L. F., & Nichols, B. E. (2021). Population density and basic reproductive number of COVID-19 across United States counties. *PLoS One*, 16.
- WTOP News. (2020). Dr. Fauci: Protests are 'perfect setup' for coronavirus spread, preprint, infectious diseases (except HIV/AIDS).
- Yanagizawa-Drott, D., & Madestam, A. (2011). Shaping the nation: The effect of fourth of july on political preferences and behavior in the United States. *SSRN Electronic Journal*.
- Dave, D.M, A.I Friedson, K. Matsuzawa, J.J Sabia, and S. Safford, 2023, black lives matter protests and risk avoidance: The case of civil unrest during a pandemic, *Journal of Human Resources*, preprint at <https://doi.org/10.3368/jhr.0121-11463R1>.