School Reopenings, Mobility, and COVID-19 Spread: Evidence from Texas

Charles Courtemanche*
University of Kentucky, NBER, and IZA

Anh Le
University of Kentucky

Aaron Yelowitz
University of Kentucky and IZA

Ron Zimmer
University of Kentucky

August 2022

Abstract: In this study, we examine the effects of fall 2020 school reopenings in Texas—where schools largely opened on time and during high rates of community spread—on COVID-19 cases and fatalities as well as mobility patterns. Using event-study models and county-level data, we first find evidence that reopening Texas schools gradually but substantially accelerated the community spread of COVID-19 and increased the number of fatalities. However, these results need not be solely attributable to spread within schools, as spillover effects on adult behaviors are also possible, either from the return of free child care or the signal being sent that returning to normal activities was safe. We therefore next examine changes in adult mobility using census block group-level data from SafeGraph. Event-study results show that school reopenings increased the amount of time adults spent out of the home as well as at full- and part-time work. The results are concentrated on weekdays rather than weekends, implying that free child care is a more important mechanism than signaling. Our analyses suggest that it is important to take account the general equilibrium effects of school openings, including adult behavior effects.

* Corresponding author, courtemanche@uky.edu, Department of Economics, Gatton Business & Economics Building, University of Kentucky, Lexington, KY, 40506-0034

We thank SafeGraph, Inc. for making their data available for research. We also thank Jessica Thomas, Hunter McCormick, Ben Scott, and Yaxiang Song for collecting school districts start dates and modality of instruction from web searchers. Finally, we thank Dan Goldhaber, Joe Sabia, Richard Buddin, and seminar participants at the Frank Batten School of Leadership and Public Policy and the School of Education and Human Development at the University of Virginia for helpful comments.

JEL Codes: I18, I28,
Keywords: COVID-19, coronavirus, Texas, school reopenings, mobility

I. Introduction

As COVID-19 spread across the country in 2020, school administrators and policymakers made the difficult decision of whether to reopen primary and secondary schools for in-person instruction. In the spring of 2020, nearly every state elected to close schools for in-person instruction and switch to remote learning (Kaufman et al. 2021). However, by the fall of 2020, policies varied widely, with some states and districts electing to delay school openings for in-person instruction while others opened for in-person instruction with mitigation policies in place (Harris et al., 2021). Still others largely opened schools for in-person instruction without delays and with inconsistent use of mitigation strategies such as masking and social distancing.

Initially these decisions were made with little to no evidence on the effect school openings had on health outcomes. Only recently has research emerged, largely finding that school openings did not lead to increased COVID-19 spread and hospitalization (Harris, et al., 2021; Ertem, et al., 2021; España, et al., 2021; Bosslet, et al., 2022; Goldhaber et al., 2022). Some of this research did, however, suggest that school openings could lead to COVID-19 spread in communities in which there were preexisting high rates of spread (Harris, et al., 2021; Goldhaber et al., 2022). While this prior research is informative, it has not considered the impact school openings had on community behavior more broadly and the role this may have played in the spread of COVID-19 and fatalities.

In this study, we first examine whether school openings in the fall of 2020 to in-person instruction in Texas—where schools largely opened on time and during high rates of community spread—lead to an increase in COVID-19 cases and fatalities. Using event-study models, we find evidence that reopening Texas schools gradually but substantially accelerated the community spread of COVID-19 as well as fatalities. Results are broadly similar across a wide range of robustness checks, including those that address newly discovered issues with staggered-treatment-time two-way-fixed-effects research designs.

Motivated by the large effect sizes, we then use census block group-level data from SafeGraph (which tracks the movement of individuals aged 16 and older by using cell phone data) to explore whether the effects could be at least partly attributable to broad behavior changes in the community, as opposed to solely in-school spread. Such spillovers are plausible for multiple reasons. First, reopening schools greatly reduces the cost of returning to in-person work and outside-the-home leisure activities for parents and other caregivers. To the extent that these activities increase social interactions, they could contribute to the spread of COVID-19. Second, reopening schools also sends a signal to the community that returning to normal activities is safe, which could plausibly also expedite return to work and other activities. This is similar to the "learning by deregulation" concept described in Glaeser et al. (2020). Using event-study models, we show that time spent outside the home by adults rose sharply with school openings, as did the amounts of part-time and full-time in-person work. To help distinguish between the two possible mechanisms, we split the sample into weekdays and weekends. The "reduced cost of child care" mechanism should only matter on weekdays, when school is in session, whereas the signaling mechanism would presumably matter on all days. We find that the effects on adult mobility are concentrated on weekdays, implying that they are driven primarily by reduced child care costs. These findings are important as the highlight the importance of considering the general equilibrium effects of school openings, including adult behavior effects.

II. Background

<u>*School Reopenings in Texas*</u>

On July 7, 2020, the Texas Education Agency (TEA) issued school reopening guidelines, which covered topics such as COVID-19 prevention, responses, mitigation, and information dissemination.[1] These guidelines covered the wearing of masks, reporting of positive cases, and screening of staff, teachers, and students. Most importantly, it provided the following guidance for reopening schools: “during a period up to the first four weeks of school, which can be extended by an additional four weeks by vote of the school board, school systems may temporarily limit access to on-campus instruction.”

These instructions were further clarified by a July 17, 2020 joint statement from Governor Greg Abbot, Lt. Governor Dan Patrick, Speaker Dennis Bonnen, Senate Education Chairman Larry Taylor, and House Education Chairman Dan Huberty. They stated that local school districts have the constitutional authority to decide when and how schools safely open and noted that local school boards have the authority to set the start date which could be in in “August, September, or even later.”[2] They also noted that local school boards can make these decisions “on advice and recommendations by local public health authorities but are not bound by those recommendations.” Importantly, the statement also clarified that not only could school districts start the first four weeks as a “back to school transition” with remote instruction, but school districts could extend their back-to-school transition an additional four weeks with a vote of the school board and a waiver from the state. After eight weeks, school districts could ask for an addition extension as the result of health concerns related to COVID-19 and the TEA will decide those requests on a case-by-case basis. Finally, the guidance from TEA noted that school districts must provide the option for families of remote instruction, even if the school district provides in-person instruction.

[1] https://www.wfaa.com/article/news/education/texas-students-must-wear-face-masks-at-school-tea-says/287-e2ef67ef-6ec7-4827-9a80-43fb83932564

[2] https://gov.texas.gov/news/post/governor-abbott-lt-governor-patrick-speaker-bonnen-chairman-taylor-chairman-huberty-release-statement-on-school-re-openings

However, because of the challenges of the logistics of providing both in-person and remote instruction, school districts could restrict families to switching their choice of instructional modality only at the end of grading periods.

With this policy context as background, Figure 1 displays the start date of opening schools for in-person instruction for school districts in the 2020-21 school year relative to the start date of opening schools in the 2019-20 school year.[3] About two-thirds of school districts opened schools in 2020-21 within one week of the start date of 2019-20 in spite of the widely documented surge in COVID-19 cases in Texas in the summer of 2020. Moreover, less than two percent of school districts delayed the reopening by more than eight weeks, possibly because of the requirements imposed by the state to obtain an exemption to remain virtual longer than eight weeks. To the extent that state directives trumped local caseloads or politics in influencing reopening decisions, that would help to alleviate endogeneity concerns in our econometric analysis.

Econometric Evidence on Schools and COVID-19

As the pandemic began to unfold during the spring of 2020, very little was known about the likelihood of spread among young populations and whether schools could safely operate with in-person instruction. Three early studies that controlled for other accompanying restrictions like restaurant closures and shelter-in-place orders did not find evidence that school closings slowed the spread of COVID-19 (Courtemanche et al., 2020; Hsiang et al., 2020; Flaxman et al., 2020). However, a fourth study that did not control for these other restrictions did find evidence of a sizeable effect (Auger et al., 2020).[4] These prior studies were of limited usefulness for reopening

[3] In most districts, we were able to determine the 2019-20 start date. However, in the cases where we were not able to identify the 2019-20 start date, we either used the prior year start date (e.g., 2018-19) or the median 2019-20 start date within the county.
[4] Another study (Gupta, et al., 2021) examined the early period of April and July 2020 but was not published until 2021. The study found that school openings were associated with greater levels of hospitalizations and deaths.

decisions in fall 2020 as almost all the spring school closures in the United States occurred within one week of each other, leading to little identifying variation and generally imprecise estimates. While controlling for other types of restrictions is important for causal inference, it further strains the available identifying variation, perhaps explaining the null findings from studies that did so. Further, it is not clear that closings and openings should have symmetric effects. Much more was known about mitigation strategies in fall 2020 compared to spring, but community spread was also much greater in the fall.

Only recently has econometric evidence on reopening schools begun to emerge. Isphording et al. (2020) leveraged variation in the timing of school start dates and found little evidence of effects on community spread in Germany.[5] However, the relevance of this finding for a U.S. population with different attitudes toward COVID-19 and different mitigation policies, both inside and outside of schools, is unclear.

In Indiana, Bosslet and colleagues (2021) used a panel data regression model with county and day fixed effects and examined the relationship between the proportion of students attending schools within a county in person and daily new COVID-19 cases per 100,000 residents in a county. To measure the effect, the authors did not measure the contemporaneous period health outcomes but built in a 28-day lag. While this approach allows for a period of incubation and recording of a positive test, it is not as flexible as an event study model which provides the ability to see how health outcomes changes over time after the initial treatment. The authors found a positive and significant effect. The magnitude of the effect is an additional 0.336 cases per 100,000 residents for every additional 10 percent of students attending in-person instruction. Overall, this estimate could be interpreted as a small impact. However, it should be noted that the rate of

[5] This result is consistent with two descriptive studies of small sets of schools in France and Helsinki that also found little evidence of spread (Dub et al., 2020; Fontanet et al., 2020).

community spread in Indiana was relatively small, with only about 10-20 daily cases per 100,000 residents. This is contrast to Texas in which many school districts opened with as many as 30 daily cases per 100,000 residents. Bosslet et al. (2021) also included only a six-week period for analysis and did not extend beyond Oct 6. 2020, before much of the surge of the late fall 2020, which may reduce the inferences to periods in which there is greater spread.

Tulane researchers used national insurance claims data and U.S. Department of Health and Human Services (HHS) hospitalization data along with national data on school reopenings to examine the impact of school reopenings on hospitalizations (Harris et al., 2021). Overall, they found no association between school reopenings and hospitalizations. However, they noted that in areas with higher pre-opening COVID-19 hospitalization rates, the results were less conclusive with some evidence indicating that in these areas, school openings could lead to greater hospitalizations. In their analysis, the sample period only allows for six weeks of post-treatment data, which may not be enough time for meaningful increases in hospitalizations to occur given incubation periods and the potential need for multiple rounds of spread outward from schools before reaching the vulnerable individuals who are most likely to require hospitalization.

Another study, released by a research consortium named CALDER, examined monthly county-level COVID-19 cases using school reopening information provided by Michigan and Washington's departments of education (Goldhaber et al., 2021). The researchers noted that in Washington, only 10 percent of districts (almost entirely rural) and only 2 percent of the student population was attending either a school operating with hybrid or in-person instruction. In Michigan, the percentages were higher, with 76 percent of schools operating either with hybrid or in-person instruction. Like the Tulane study, this study examined COVID-19 cases prior to much of the surge of cases in the winter of 2020-21. The research team found that in-person modality options are not associated with increased spread of COVID-19 at low levels of pre-existing COVID-19 cases but did find that cases increase at moderate to high pre-existing COVID-19 rates.

Finally, Bravata et al. (2021) used SafeGraph data merged with medial insurance claim data to examine the relationship between increased school visits (as tracked by SafeGraph cell phone data) and COVID-19. In their analysis, the authors compared families with children to families without children and found that increased school visits led to a small increase in infections. However, importantly, the authors found that the increased school visits were associated with lower rates of transmission during the initial months of the pandemic when COVID-19 prevalence was lower and that nearly all of the main effect is driven by higher transmission rates during the later months during higher levels of transmission. This finding reinforces the insights from the CALDER and Tulane studies, which raised the possibility that COVID-19 spread could occur when school opened in communities of preexisting levels of high transmission. It should be noted that defining treatment as visiting schools for families with children assumes that opening schools will not affect transmission through any other means other than within-school transmission. As discussed above, we will explore the possibility that the opening of schools may cause adults to be more mobile, which could lead to transmission outside of schools that is still causally linked to their opening.

Our work complements these other studies in multiple ways. First, Texas gives us a glimpse into a state that opened relatively normally as most school districts opened on time. In addition, while many states opened schools using hybrid models where only partial numbers of students attended schools each day to allow for greater social distancing, most schools in Texas opened at near capacity. Specifically, in reviewing school opening plans of Texas school districts, our best estimate is that around 90 percent of school districts opened fully in-person without any staggered or phased-in attendance. This is in contrast to 42 percent nationally (Harris et al., 2021). Also, masks were not required in most Texas schools. Therefore, our analysis in Texas provides a stronger glimpse into the effects of current policies school districts are adopting across the country, which relies less upon hybrid models or mask mandates. Second, studying Texas allows us to examine the opening of schools during relatively high community transmission as Texas had higher rates of COVID-19 cases than the national average through August and September. Third, and finally, much of the current research has not considered the effects school openings had on the broader communities' behaviors. In our analysis, we examine the effect school openings had on adult mobility by tracking cell phone data. In this way, we are able to provide insights into any the mechanisms behind any possible adverse health effects we observe in our analysis.

III. Data

To collect information for each school district's start date and modality, we performed Google searches in which a team of assistants searched for key terms using district name and the phrase "back to school plan".[6] The vast majority of districts had a back-to-school plan and it often included both the district's modality plan for instruction and the school district start date. If the school district started with virtual instruction, the back-to-school plan often listed the planned date for in-person instruction.[7] In cases in which the start date was not listed, the team of assistants searched for the school district's academic calendar. In cases where back-to-school plans or calendars were not available, we also conducted newspaper and Facebook searches to identify this

[6] Like the Tulane study, we did not include charter or private schools primarily because they represent a small minority of the total students in the states and also because it would have been difficult to ascertain this information.
[7] In some cases, districts phased in in-person attendance (e.g., Kindergarten through 3rd grade could attend in person one week and the following week the rest of the grades could attend in person). In these cases, we used the first date students were allowed on campus. If the district only allowed special education students on campus, we did not count this as in-person instruction given the small number of students on campus.

information through news stories and school district's Facebook posts.[8] Even in cases where a back-to-school plan and/or academic calendars were available, we often conducted additional newspaper or Facebook searches to verify the district's start date and modality of instruction.

Because COVID-19 cases and fatalities are only available at the county level, we need to aggregate the school reopening variable from the district to the county level, which requires accounting for the fact that not all districts within a county opened at the same time. In the Tulane study, the researchers defined treatment as occurring when the *first* district within a county reopened. However, for many districts in Texas, this definition would result in a county being labeled "treated" when only a small fraction of schools is actually open. Consider Bexar County, a large county that includes San Antonio. Southwest Independent School District (ISD), which represents less than 5 percent of the county's student enrollment, was the first district to open schools on August 24, 2020. However, there were some districts within the county that opened up schools as late as seven weeks later and six school districts representing 75 percent of the county's student population opened on September 8, 2020. In this case, defining treatment based on the earliest opening school district would effectively lead to it being assigned two weeks too early relative to when the majority of students within the county began in-person instruction. Therefore, our primary treatment definition is the week in which 50 percent of the county's students attended

[8] After these steps, there were only 11 school districts in which we could not identify the start date and only 17 school districts we could not identify the modality of instruction. We tried to follow up with each district with a phone call. Through these phone calls, we were able to identify the start date for seven of the 11 missing dates for school districts and the missing modality information for 12 of the 17 school districts. Therefore, we had missing dates for four school districts, which we imputed based on the median start date within their county. For modality, we had missing dates of five school districts, which we imputed as the majority instructional modality of the school districts within the county. These are very small districts, with the average size of the missing start date districts being 78 students and the average size of the districts with missing modality information being 177 students. Since our data will be population-weighted, these districts have little consequence for the results.

schools that had begun in-person instruction. In the case of Bexar County, that would be the week of September 8th.[9]

We should also note that treatment begins for our empirical analysis once schools open for any type of in-person instruction including fully in-person, phased-in (e.g., a subset of grades open for in-person instruction with gradual number of grades eligible to attend in-person over time), or as a hybrid model (e.g., students attending in person part of the week and attending virtually the rest of the week). However, as discussed previously, phased-in and hybrid reopenings were rare in Texas.

In opening schools for in-person instruction, districts almost uniformly allowed families to choose to attend in-person or remotely. However, districts had to prepare for the possibility that all or nearly all students could attend in person. Therefore, our main treatment variable could be thought of as "intent-to-treat" (ITT) analysis as a district's decision to provide in-person instruction is providing the opportunity for all students to attend in person. That said, schools in Texas tended to open at relatively close to full capacity as nearly 60 percent of all school districts had 80 percent or more of their students enrolled for in-person instruction by the end of September. Our analysis is also an ITT analysis in a second way. Once treatment begins by a school district opening schools for in-person instruction, we consider the school opened throughout the analysis, even if the school has a temporary shutdown as a result of an outbreak. In defining treatment in this way, our estimates should be seen as conservative.

Our COVID-19 data come from the Texas Department of State Health Services (TDSHS).[10] Numbers of COVID-19 cases, fatalities, and tests are recorded daily at the county

[9] Later, we present a series of analyses that suggests our results are robust to alternative definitions of treatment including using the first school district that opened schools in person.
[10] https://dshs.texas.gov/coronavirus/additionaldata.aspx

level from May 3, 2020 through January 3, 2021. We use weekly (Sunday through Saturday) data instead of daily data because not all labs are open daily or do not report daily (e.g., many labs are not open on weekends) and can have duplicate numbers or reporting errors, which can lead to oscillating numbers from one day to the next. By using weekly numbers, we are largely able to smooth out these fluctuations.[11] To account for variations in county population, we calculated COVID-19 cases, fatalities, and tests per 100,000 residents using 2019 county population estimates from the Census Bureau.[12] These cases and fatalities variables are our main outcome variables, while the testing variable is a control in the cases regressions.

Importantly, no binary treatment definition perfectly captures treatment timing in Texas given the often-staggered nature of reopenings within counties. Only 89 of Texas' 253 counties reopened all at once. In 93 of the remaining 164, the first school district in the county opened one week before enough others opened to push the county across the 50% threshold. In the other 71 counties, the lag between first reopening and crossing the 50% threshold was longer, reaching up to eight weeks in two cases. Across the entire state, the population-weighted average length of time between first opening and crossing the 50% mark was 2.49 weeks. The econometric implication of this somewhat fuzzy treatment design is that effects could plausibly occur prior to our officially designated treatment date. Effects that emerge one or two weeks before treatment could potentially be causal rather than driven by problematic pre-treatment trends. This is more probable for the mobility outcomes, which should be affected contemporaneously, than the COVID-19 cases and deaths outcomes, for which a lag before impact is likely.

---

[11] It should be noted that some data errors within the TDSHS data systems have been discovered over time as documented by media accounts: https://www.khou.com/article/news/health/coronavirus/texass-record-high-covid-positivity-rate-falls-after-data-experts-investigate/287-ffc19167-0d47-4be9-8c06-8648229288ef and https://www.texastribune.org/2020/09/24/texas-coronavirus-response-data/. Corrections to these errors could cause accumulated cases or tests to decrease over time as the data are corrected. These anomalies should create noise, but not bias and should largely be accounted for in our analysis using week fixed effects.

[12] https://www.census.gov/data/datasets/time-series/demo/popest/2010s-counties-total.html

To help understand potential spillover effects of school reopenings on adult mobility, we utilize Social Distancing Metrics (Version 2.1, "SDM") data provided by SafeGraph, Inc., from May 3, 2020 to January 3, 2021.[13] SafeGraph collects information on almost 45 million cellular phone users, including about 10 percent of devices in the U.S. The sample correlates very highly with the true Census populations with respect to distribution by county, educational attainment, and income.[14] These data are aggregated from GPS pings provided by cellular devices that have opted-in to location sharing services from smartphone applications. The device data is aggregated by Census Block Group (CBG) and day, based on a device's "home" location.[15] In our timeframe, there were 15,705 CBG's overall in the Texas SDM; on an average day, more than 1.9 million devices were followed in Texas. For our analysis, we restricted the sample to a balanced panel of 14,580 CBG's (with more than 1.6 million overall devices on an average day).[16] The typical CBG had approximately 112 devices. We created samples at the weekly level for the full week (Monday through Sunday), for weekdays (Monday through Friday), and for weekends (Saturday and Sunday).

We utilize four of the mobility measures provided in the SDM that are often used in other studies. The most common measure is the fraction of devices that do not leave their home location

[13] https://www.safegraph.com/blog/stopping-covid-19-with-new-social-distancing-dataset
[14] https://www.safegraph.com/blog/what-about-bias-in-the-safegraph-dataset
[15] To impute a "home" location for a cellphone user, SafeGraph considers a common nighttime location of each mobile device. In the entire United States, the SDM is aggregated to approximately 220,000 CBGs. To enhance privacy, CBG's are excluded if they have fewer than five devices observed in a month.
[16] CBG's were excluded if (a) the CBG was not observed for all days in our sample period, (b) the CBG could not be merged to demographic information from the 2018 American Community Survey (ACS) 5-year estimates, (c) the CBG's population – according to the 2018 ACS – was in the bottom or top 1 percent of the full distribution (corresponding to 391 and 7150, respectively), or (d) over the course of the panel, relative to the mean device count in the CBG, any specific CBG-day observation had a device count that more than twice the mean or less than half the mean. By restricting to CBG's with relatively stable numbers of devices over the long panel, we hope to avoid complications related to installation and removal of apps, inactive devices, and sample attrition highlighted in some other studies (Andersen et al., 2020; Allcott et al., 2020). Although Safegraph reports that some apps implement GPS collection methods that depend on the movement of the device (rather than a fixed time interval), this would likely affect levels of certain metrics (e.g., completely home all day) but not changes.

during a given day ("Percent Completely Home").[17] We also use two "work" measures. SafeGraph defines "work" as either the fraction of devices that spent more than 6 hours at a non-home location between 8am-6pm ("Percent Full Time") or fraction of devices that spent between 4-6 hours at a non-home location between 8am-6pm ("Percent Part Time").[18] Finally, several studies have examined median time spent away from home (or at home).[19] These measures are based on the observed minutes outside of home (or at home) throughout the day, regardless of whether these time episodes are contiguous. The time during which a smartphone is turned off is not counted towards the measures.

Finally, for some of our analyses, we utilize county-level variables from other sources. The county's college enrollment is available from the U.S. Department of Education National Center for Educational Statistics (NCES).[20] Percent of voters who voted for President Trump in the 2016 presidential election comes from the MIT Election and Data Science Lab (2018). We control for average weekly temperature, precipitation, and snowfall using data collected by the National Oceanic and Atmospheric Administration (NOAA) and the Global Historical Climatology Network.

[17] See Bailey et al. (2020), Bullinger et al. (forthcoming), Cronin and Evans (2020), Allcott et al. (2020), Dave et al. (2020a), Simonov et al. (2020), Dave et al. (2021), Friedson et al. (Forthcoming), and Gupta et al. (2020).

[18] See Bullinger et al. (forthcoming) and Simonov et al. (2020).

[19] See Allcott et al. (2020), Dave et al. (2020a), Cotti et al. (forthcoming), and Gupta et al. (2020).

[20] These data were collected at http://nces.ed.gov/ccd/elsi/.The reporting years of enrollment ranged from 2013-2017. As part of the data cleaning process, for residential campuses only, we assumed all enrolled students could attend classes in person and therefore, we calculated the maximum weekly proportion of the total county population that could be on campus by dividing the number of enrolled students by the county population. To calculate the daily proportion of college students of the total county population, we assumed that no students were on campus during the summer (nearly all colleges did online instruction over the summer). We also assumed all residential colleges had in-person classes for the fall semester. For those colleges with no residential students, we assumed the colleges were providing instruction either online or had minimal student interactions. Using Google searches of academic calendars, we identified the start date for each college, which is the day we assumed students began interacting on campus. In many counties, there are multiple colleges with different start dates, which means the college proportion changes over time as more and more colleges start their fall sessions.

Our main analysis sample contains a balanced event-time window surrounding treatment, i.e. the week of the county's largest increase in percentage of students who can attend in-person school. For the COVID-19 outcomes, we include eight weeks prior to treatment, the treatment week, and eight weeks after treatment. A lengthy post-treatment period allows for multiple rounds of spread (e.g. from student to parent to grandparent), incubation periods, time to receive and obtain results from a test, and the fact that deaths can occur weeks after infection. On the other hand, a long post-treatment period faces a relatively high risk of confounding from other concurrent shocks. In our case, the holiday break – which started in many Texas districts after the week of December 13 – is a particular concern, as schools being "reopened" should not influence spread when they are not in session. In our view, an eight-week post-treatment window best balances these considerations. It is long enough to plausibly capture much of the dynamics of the treatment effect. At the same time, it is short enough to avoid sample windows that stretch past the week of December 13 for all but two small counties (Starr and Zavala) that will have little influence in our population-weighted sample. For the SafeGraph mobility outcomes, there is not a clear reason to expect a lag before treatment effects emerge, so we limit the event-time window to six weeks on each side, thereby ensuring that the sample window does not extend past the week of December 13 for any county.

Table 1 shows means and standard deviations for our outcome variables in both the pre- and post-treatment periods, weighted by population. Interestingly, new cases per capita were about the same in the pre- and post-treatment periods, while death rates went down by almost 50 percent. This was in spite of a moderate increase in mobility across all four measures. Of course, numerous factors affect these flat or downward trends, including better understanding of preventive measures such as mask-wearing, advancements in treatments, and the average age of cases gradually becoming younger. A finer-grained econometric analysis is necessary to disentangle the causal effects of school reopenings from these underlying trends.

Table 2 shows results from a simple cross-sectional regression of week of reopening (ranging from 14 to 28, with week 1 being the week of May 3) on several county-level variables that might be expected to influence reopening decisions: President Trump's 2016 vote share, percent Hispanic, percent Black, county population, and percent of the SafeGraph sample who stayed completely at home for the day in the four weeks prior to any schools reopening (a proxy for compliance with public health guidelines), and average weekly new cases per capita in the four weeks prior to any schools reopening. We standardize the covariates to allow a direct interpretation of the magnitudes. Trump vote share is the dominant predictor, which is consistent with previous research that showed politics drove school opening decisions (Valant, 2020). Each standard deviation increase in Trump vote share is associated with schools reopening 1.2 weeks sooner. In contrast, none of the other variables are statistically significant, and none have a magnitude greater than 0.19 weeks. The coefficient for pre-school-year caseloads is small and highly insignificant. Therefore, reopening decisions appear to have been driven much more heavily by politics than public health considerations, which may be surprising but is consistent with prior research (Valant, 2020). This can be seen as favorable for an econometric analysis, as it suggests that reverse causality from caseloads influencing reopening decisions should not be a concern. We will be able to account for stable county characteristics such as political views by including county fixed effects.

IV. Econometric Methods

We aim to identify the causal effects of school reopenings on new weekly COVID-19 cases and fatalities per 100,000 residents by estimating event-study regression models of the form

$$y_{ct} = \beta_0 + \sum_{i=-8, i \neq -1}^{8} \beta_{1i}\, OPEN_{c,t-i} + \beta_2\, TESTS_{ct} + \alpha_c + \tau_t + \varepsilon_{ct} \qquad (1)$$

where the subscripts c and t represent county and week; y is the case or fatality outcome; $OPEN$ is the reopening indicator; $TESTS$ is a control variable for the number of COVID-19 tests per 100,000 residents,[21] included since differential testing rates across locations and time can be an important driver of confirmed case numbers; α and τ are county and time fixed effects; and ε is the error term. Observations are weighted by county population, and standard errors are robust to heteroskedasticity and clustered by county.

The summation term for the treatment variable reflects the inclusion of separate indicator variables for whether schools will reopen eight weeks after week *t*, seven weeks after, six weeks after, etc.; whether schools reopened exactly in week *t*; and whether schools reopened one week before week *t*, two weeks before, etc., up to eight weeks before. The variable for whether schools will reopen in t-1 is omitted as the reference period. The "lead" terms (weeks until school reopening) measure pre-treatment trends, while the "lag" terms (weeks after school reopening) measure the evolution of the treatment effects over time. As discussed above, we expect the effects on new cases to grow over time because of the incubation period, the lag between symptom onset and receiving a test, the time required to obtain test results, and the exponential nature of case growth. For fatalities, we expect an even longer lag since deaths typically occur after an extended battle with the illness.

We also estimate a number of variants of our baseline event-study specification as robustness checks. The first two robustness checks use alternative definition of treatment. First,

[21] Since test results might not be recorded in the same week that the test was conducted, we experimented with including lags of the testing variable, finding that the contemporaneous value as well as two weekly lags were statistically significant. We therefore include all three of those variables in the regressions.

we drop 57 counties that were already ≥ 40% reopened prior to the treatment week. Since we define treatment as being 50% reopened, a county that was already 40% reopened might not experience much of a jump during its "treatment week", leading to substantial measurement error. Second, we define treatment as the week during which the county had its first district reopening, which is consistent with the treatment definition used by the Tulane study

The next two checks add variables in an effort to address possible omitted variable bias concerns. Causal inference in our event-study model requires the assumption that case and death trajectories would have evolved similarly in early versus late reopening counties in the counterfactual in which schools did not reopen. The pre-treatment trends estimated using the lead terms in the event-study model are informative as to how case and death trajectories would have evolved in the counterfactual scenario. However, it is possible that some confounders did not emerge until the post-treatment period. For instance, most Texas colleges and universities opened for in-person instruction at the start of the fall semester. If these post-secondary reopenings fueled COVID-19 spread and if school reopening dates were also systematically correlated with the prevalence of college students in the county, this could bias our estimators for the school reopening coefficients. We therefore estimate a model that controls for college and university reopenings in a dose-response, event-study manner. Specifically, we construct a variable for the proportion of a county's population that attends an in-session post-secondary institution in a given week. We then interact this continuous "dosage" measure with indicators for each of the eight weeks before and after the first college reopening in the county. For our second check in this category, recall that the results from Table 2 showed that vote share for President Trump was the dominant predictor of reopening week. Residents' political views are presumably fixed during a two-month sample period, meaning that they are captured by the county fixed effects. However, it is possible that political views could influence not only *levels* of new COVID-19 cases but also *trends*, and county fixed effects alone would not account for the latter. If heavily Republican counties opened schools relatively early and also developed steeper COVID-19 trajectories in the fall for reasons besides school reopenings, our estimated effects of reopenings would be biased upwards. We therefore estimate a model that adds interactions of time-invariant Trump vote share (coded as dummies for each of the four quartiles) with each week fixed effect, thereby flexibly allowing for right- and left-leaning counties to have different COVID-19 trajectories.

The next robustness check shortens the sample window from eight weeks on each side of treatment to six. If the results vary considerably with simple changes in sample bandwidth, this would be suggestive of deeper specification problems.

Finally, an emerging literature documents problems with two-way fixed-effects (TWFE) models with staggered treatment times.[22] First, TWFE regressions give more weight to observations treated in the middle of the sample period, which can lead to unreliable estimates of the average treatment effect if treatment effects are heterogeneous. Using the event-study formulation with a balanced panel and a sample period centered around treatment time rather than calendar time alleviates this concern. Since each county has exactly eight pre-treatment observations, one observation during the treatment week, and exactly eight post-treatment observations, the variance of each treatment variable is identical for each county.

More troublesome in our context is that, in settings that rely exclusively on variation in treatment timing for identification as opposed to having control units, two-way fixed effects models implicitly use early treated units as controls for later treated units. This leads to bias when

[22] This literature includes Callaway and Sant' Anna (forthcoming), de Chaisemartin and D'Haultfoeuille (2020), Goodman-Bacon (forthcoming), and Sun and Abraham (2020). Our discussion in the remainder of this section is based on reviews of this emerging literature by Baker et al. (2021) and Cunningham (2021, pp. 461-510).

treatment effects are dynamic because the response of the early treated units is still evolving at the time that they are called upon to be controls, effectively leading to a violation of the parallel trends assumption for those particular late-versus-early comparisons. Event-study models do not necessarily alleviate this concern. Under the assumption that the treatment effect either strengthens or stays the same over time, the bias is toward zero and we can conclude that, if anything, our estimates are conservative. We find this assumption plausible for COVID-19 outcomes; as discussed above, all the reasons to expect treatment effects to evolve over time point towards them becoming stronger rather than weaker.

Nonetheless, we conduct two robustness checks that utilize newly developed methods that address this issue. Both of these methods perform well in simulations and applications conducted by Baker et al. (2021). First, we employ the "stacked regression" strategy used by Cengiz et al.'s (2019) study of four decades of state minimum wage increases. This method begins by constructing new datasets for each treatment event (each county's school reopening) along with corresponding "clean controls", defined as those counties whose school reopenings did *not* occur within eight weeks on either side of the reopening week of the focal county. Then, we combine the resulting datasets into a single "stacked" sample and re-run the baseline regression, except adding interactions of indicators for each underlying dataset with each of the county and week fixed effects (as well as, when COVID-19 cases is the outcome, the testing controls). Standard errors are clustered by county to prevent the duplication of data from leading to over-rejection of the null hypotheses. Our other robustness check implements the method of Callaway and Sant' Anna (forthcoming), which first estimates dynamic treatment effects for units treated at each time period, then combines them by weighting by sample share rather than treatment variance. This method also purges the potentially problematic late-treated versus early-treated-as-control comparisons from the identifying variation.[23,24]

V. Results

Figure 2 displays the event-study results for the baseline model with new COVID-19 cases per 100,000 residents as the outcome. The dots indicate the coefficient estimates for each week of event time relative to the reference period of one week before reopening. The bars represent 95 percent confidence intervals. As a point of reference for evaluating magnitudes, recall from Table 1 that the pre-treatment sample mean for the dependent variable is 147.7 cases per 100,000. The solid vertical line represents treatment time. The dashed vertical line denotes the 2+ week average "lead time", i.e. the time between first reopening in the county and when it crosses the 50% threshold and therefore is officially classified as treated. Accordingly, effects in weeks -2 or -1 are not necessarily indicative of problematic pre-trends.

The results provide evidence of a positive, large, and causally interpretable effect of reopening schools on COVID-19 cases per 100,000 residents. The coefficient estimates associated with the negative event time terms show little evidence of problematic pre-treatment trends, including in the ambiguous two weeks before treatment. The line is nearly straight, and the point estimates are all small and never close to statistically significant at the 5 percent level relative to

---

[23] To implement this method, we use the open-source STATA and R packages provided by Jonathan Roth and Pedro Sant'Anna (footnote: https://github.com/jonathandroth/staggered#stata-implementation). For COVID-19 cases, the method requires us to drop three counties that are the only county treated in a particular week. For fatalities, we encounter a problem with singular variance matrix because small counties tend to have weeks in which there were zero deaths reported. We therefore limit the sample to counties with more than 19,000 residents and shorten the event study window to seven periods before and after reopening to avoid unbalanced treatment groups.

[24] Note that we do not also present results from the Goodman-Bacon (forthcoming) decomposition because that is designed for two-way fixed effects models with a single treatment variable, rather than for event-study models like ours with numerous treatment variables. That said, if we run a basic TWFE regression with a single treatment variable, the decomposition shows that the treatment effect estimate is driven roughly equally by early-treatment versus late-treated-as-control and late-treatment versus early-treated-as-control comparisons. The estimated treatment effect from the former is more strongly positive than that from the latter, consistent with dynamic treatment effects causing a bias toward zero when early-treated units are used as controls, as discussed above.

the reference period. The coefficient estimates from the post-treatment period show that a statistically significant increase in cases which generally increases over time reaching a peak of around 140 new cases per 100,000 residents. This effect size is substantial, as it is nearly equivalent to the pre-treatment sample mean. The later period confidence intervals are large, but even the low end of the 95 percent confidence interval for the week eight coefficient estimate would represent about a 50 percent increase relative to the pre-treatment mean.

The results from the robustness checks for new cases, shown in Appendix Figures 1-6, are broadly similar. In all regressions, the estimated effect of reopening schools is positive, with a general pattern of strengthening over time (although individual coefficient estimates sometimes deviate from that pattern).[25] The point estimate in the last week is never below 110 cases per 100,000, meaning that the estimated effects are consistently large. Of special note is the robustness check in Appendix Figure 2, which uses as an alternative definition of treatment the first school district opening within a county rather than when 50 percent of students are eligible to return to school. Comparing Figure 2 and Appendix Figure 2, there is a noticeable difference in the timing of the jump in cases: it occurs about two weeks later when using the treatment definition of the first school district opening as opposed to when 50 percent of students within the county can enroll. This difference is consistent with the often substantial lag between when the first school district in a county opens for in-person instruction and when the county reaches the 50 percent threshold.

Moving to our analysis of fatalities, Figure 3 shows the baseline results for weekly deaths per 100,000 residents, which has a pre-treatment mean of 3.51. As with cases, the results suggest a positive causal effect of school reopenings. During the pretreatment period, there is a negative trend (although they are never statistically significant) for week eight through week five prior to

[25] Also, in the online appendix Figures A1-A3, we show additional robustness checks that show the same general patterns of growth in COVID 19 cases over time.

treatment, which corresponds to the relative mild period of COVID spread of late spring and early summer of 2020. For remaining periods of pretreatment, the estimates are statistically insignificant and flat. A small statistically significant increase in deaths emerges the week following reopening. This could plausibly represent delayed effects of the school districts that opened in the weeks prior to our official definition of treatment. The effect becomes slightly insignificant in the next week, before becoming significant again in week three. It then steadily grows and remains significant in subsequent weeks, with the point estimate reaching about 2.5 deaths per 100,000 residents after eight weeks. This magnitude represents more than two-thirds of the pre-treatment sample mean, and the low end of the 95 percent confidence interval is just under 1.5, which is still substantial.

The results from the robustness checks, presented in Appendix Figures 7 through 12, are again broadly similar in terms of signs and significance. Across the models, the results are substantively consistent with the baseline results.[26] The effects are always positive and grow over time, with the last period point estimate never being below 1.3. Note that in one case, the regression with the first reopening week being coded as the treatment time, the post-treatment coefficients never quite reach statistical significance. However, this is attributable to wide confidence intervals rather than a meaningful reduction in effect size: the last period coefficient estimate is still around 3.5.

In order to help assess the practical significance of the results, we utilize the estimates from the baseline models for cases and fatalities to predict how Texas' COVID-19 trajectory would have evolved differently if schools had not reopened. As discussed above, the generally large confidence intervals associated with our estimates mean that relying exclusively on point estimates

[26] Also, in the online appendix Figures A4-A6, we show additional robustness checks that show the same general patterns of growth in COVID 19 fatalities over time.

for these calculations could be misleading. We therefore also perform a more conservative simulation using the low end of the estimates' 95 percent confidence intervals.

First, we compute the predicted number of cases attributable to school reopenings. Our point estimates for reopening in the present week, the prior week, two weeks ago, and so on out to eight weeks ago, are 45.33, 23.98, 38.13, 45.10, 74.44, 66.25, 117.69, 120.14, and 143.30, respectively. After eight post-treatment weeks, the cumulative number of extra cases is the sum of all nine coefficient estimates, which is 674.36 per 100,000 residents. Since our regression is weighted by population, our estimates are interpretable as average effects across all of Texas. Therefore, the total number of extra cases is given by multiplying 674.36 by the state's population of 28,995,712 and then dividing by 100,000, yielding 195,535. According to our data, there were a total of 373,323 new cases in Texas in the nine weeks included in our post-treatment window (including the treatment week itself). Therefore, the point estimates imply that Texas' caseload would have been 52 percent lower during that time had schools not reopened.

As stated above, we caution against a literal interpretation of that number given the relatively wide confidence intervals associated with our estimates. A safer interpretation can be obtained by instead using the low end of the 95 percent confidence interval to determine the minimum number of cases attributable to school reopenings implied by our results. The low end of the 95 percent confidence intervals associated with the variables for the treatment week and each of the eight post-treatment weeks are 8.91, 2.86, 9.16, 2.85, 29.14, 15.88, 64.00, 58.51, and 74.28, for a total of 265.59. Scaling up to the population of Texas yields a *minimum* of 77,010 cases attributable to school reopenings in the nine subsequent weeks, or 21 percent of the state's total caseload during that time.

The same process can be used to compute the number of fatalities attributable to school reopenings. The baseline regression's point estimates for the treatment week and eight post-treatment week variables are 0.11, 0.31, 0.38, 0.63, 0.90, 1.13, 1.50, 1.87, and 2.52 for a total of 9.35 deaths per 100,000 residents, or 2,711 across the state of Texas. The corresponding low ends of the 95 percent confidence intervals are -0.16, -0.03, -0.07, 0.09, 0.19, 0.33, 0.69, 0.92. and 1.32, which sum to 3.28 fatalities per 100,000 residents, or 951 total across the state. During the time frame, there were 4,796 COVID-19 fatalities in Texas, so the point estimates imply that 57 percent of them were due to school reopenings, while the confidence intervals imply that *at least* 20 percent of them were.

In sum, even under conservative assumptions, reopening schools had a meaningful impact on both COVID-19 cases and associated fatalities in Texas. It is noteworthy that the percentage impacts on both outcomes are roughly similar. *Ex ante*, one might have expected the increase in deaths to be much smaller proportionally than the rise in cases. COVID-19 mortality rates are very low for children and are much smaller for the working-age adults who comprise the majority of school teachers and staff than they are for elderly or vulnerable adults. Our results therefore suggest that school-reopening-induced COVID-19 spread is reaching more vulnerable segments of the population. One possible explanation is secondary spread, where infected kids or employees spread the virus to older, more at-risk individuals. Another possibility is spillover effects, where schools opening signals to the community that it is safe to return to normal activities including returning to in-person work, leading to spread across all segments of the population that may not originate in schools. Such indirect effects could also help to explain the large effect sizes. The next section explores the possibility of spillovers more directly.

VI. Spillover Effects on Mobility

We next use SafeGraph data to explore whether changes in mobility patterns among adults may help to explain the large sizes of the effects of school reopenings on COVID-19 cases and deaths. Our baseline regression is again an event-study model given an equation like (1), with the reopening variable defined by the week in which 50 percent of the county enrollment could attend schools in person. However, we make four small changes in order to customize the approach for mobility outcomes. First, in contrast to the lags inherent in COVID-19 cases and deaths, effects on mobility can emerge immediately, and it is not obvious that they will evolve over time. Therefore, we shorten the window on each side of treatment to six weeks rather than eight, which prevents any counties' post-treatment windows from extending into the holiday break. The analysis therefore uses 13 weeks of data, and given our numbering convention, goes from -6 weeks to +6 weeks (where we denote week 0 as the week of school reopening within the county). Second, we now arrange weeks from Monday to Sunday, rather than Sunday to Saturday as we did in our models for COVID-19 spread, so that we can also examine weekday mobility separately from weekend mobility in some specifications. Third, the unit of observation is census block group (CBG)-by-week rather than county-by-week. Finally, we remove the testing controls since they are less clearly related to mobility than to the COVID-19 outcomes.

We aggregate SafeGraph's SDM database to the weekly level (averaging mobility measures across the week), where our unit of observation is a CBG. After a number of screens to the SDM data (discussed in the data section earlier), we examine 14,580 neighborhoods from 252 of the 254 Texas counties. The typical CBG in our sample has a population of approximately 1,500 people. Our four mobility measures, following SafeGraph's SDM conventions, are percentage of devices completely at home, percent part-time work, percent full-time work, and median minutes outside of the dwelling; SafeGraph's convention is to define part-time (full-time) "work" as spending 3-6 hours (6 or more hours) at one location other than home between 8 am and 6 pm local time.[27]

Using these SafeGraph definitions, in the weeks prior to reopening, approximately 28 percent of devices were completely home on a given day, and nearly 8 percent were engaged in part-time work and 4 percent in full-time work on a daily basis. In addition, the median time spent outside of the home on a given day was 108 minutes.

The event-study specifications provide evidence of increased mobility. As illustrated in Figure 4, the pre-treatment trends for all four outcomes or approximately flat. There is some evidence of anticipation effects in the week immediately prior to reopening with significant increases in work behavior. By the week after reopening, there is statistically significant evidence of increased mobility according to all four outcomes, and the effects only strengthen after that. In the first week after school reopening (week 1), there is a reduction in staying completely home of 0.75 percentage points, increases in the probability of part-time and full-time work of over 0.3 percentage points, and increases in time outside the home of almost 10 minutes. These results persist – and all grow substantially larger – in the subsequent weeks. By the end of the period, relative to the baseline of the week prior to reopening, the estimated effects are decreases of 1.3 percentage points in the probability of being completely at home all day, increases of about 0.75 percentage points in the likelihood of part and full time work, and an increase of almost 20 minutes in time outside the home. Relative to the pre-treatment sample means of the outcomes, these numbers correspond to effect sizes of approximately 5%, 10%, 19%, and 19%. While it is difficult

[27] https://docs.safegraph.com/docs/social-distancing-metrics

to credibly assess how much these changes in mobility could have contributed to COVID-19 spread, the effects size seem large enough to be consequential.

Recall that there are two primary ways in which school reopenings can lead to spillover effects on adult mobility. First, opening of schools decreases childcare responsibilities for parents and other caregivers, which could lead to either greater physical presence in workplaces or increased outside-the-home leisure activities. Second, reopening schools could send an incorrect signal to the larger community that normal activities are safe again. Such a signaling effect could even extend to those with no direct ties to students or school employees.

In an effort to distinguish between these two possibilities, we stratify the sample into weekdays and weekends. Schools operate during weekdays. Thus, increases in daily mobility induced by lower child care costs should be concentrated on weekdays. Therefore, if there are substantial effects on weekends, this would provide evidence for the signaling hypothesis. When we run the same event study models on weekdays only in Figure 5, the mobility effects are stronger than those for the full sample. By week 6, the effects on Pr(staying at home), Pr(part-time work), Pr(full-time work), and minutes of non-home time are around -1.9 percentage points, 1 percentage point, 1 percentage point, and 28 minutes. Turning to weekends only in Figure 6, we see little evidence of clear effects on any of the outcomes. Part-time work rises in the post-treatment period, but this appears to be a continuation of an upward pre-treatment trend, as opposed to a causal effect. Effects on the other three outcomes are rarely statistically significant.

Collectively, these results suggest that reopening schools leads to important spillover effects on adult mobility that may help to explain the large effect sizes for the COVID-19 outcomes. The evidence is consistent with parents going physically back to work and perhaps also increasing outside-the-home leisure activities. These effects could be due to lessened child care responsibilities, signaling about the safety of returning to normal activities, or a combination of both. However, the fact that the effects are concentrated on weekdays provides suggestive evidence that the child care mechanism is more important.

Finally, we re-examine our main mobility results with a series of robustness checks that largely mirror those for COVID-19. (These need to be added.)

VII. Conclusion

In this study, we examine the impact of opening Texas public schools for in-person instruction in fall 2020 on community spread of COVID-19 as well as fatalities. In the eight weeks after reopening, we conservatively estimate, based on lower bounds of confidence intervals, that there would have been at least 70,000 fewer COVID-19 cases and at least 950 fewer fatalities. These results hold across a variety of specifications and robustness checks. These results could be explained both by the direct effect of spread within the schools and the indirect effects of increased mobility within the community. Our analysis of cellphone data suggests that adult mobility increased along several dimensions after schools reopened, particularly on weekdays. This suggests that decision makers need to think strategically about how to encourage behavior to mitigate spread of COVID-19 not only within schools, but within the community at large.

On the surface, our empirical findings diverge with several popular narratives that have emerged about school openings. Some studies – including a prominent CDC study from Wisconsin – rely on contract tracing efforts to quantify impacts of school reopening. The imperfections of run-of-the-mill contact tracing efforts – including the inability to follow asymptomatic cases or lack of cooperation in finding all close contacts – suggests estimates of in-school spread may be a lower bound. Importantly, this approach does not account for inevitable, indirect behaviors – such as greater parental mobility including increased physical presence in the workplace – which also may contribute to community spread. Although other recent research teams (Tulane, CALDER) take methodological approaches closer to our approach and find overall more modest effects on COVID-19 spread, it is important to emphasize that the initial conditions in Texas were more ripe for community spread and schools opened more widely, more quickly, and generally, close to full capacity.

Although it is beyond the scope of our study to provide a cost-benefit analysis of school reopenings, our quantitative findings contribute a key input into such an analysis. Recent work by Kniesner and Sullivan (2020) estimate non-fatal economic losses of about $46,000 per case, and Department of Transportation apply an $11 million loss per fatality. Such health- and productivity-related losses from COVID-19 must be weighed against learning losses for children, other ancillary effects related to child mental health and abuse, and these losses could be substantial but will only become clear over time. Distributional considerations are also important, as benefits of school closures accrue disproportionately among older individuals, whereas the costs are largely borne by children. Moreover, our evidence on spillovers suggests that the value of in-person work and outside-the-home leisure activities also need to be factored into the analysis.

Obviously, as vaccinations expand, the cost-benefit calculations of opening schools changes. As of early February 2022, approximately 64 percent of the U.S. population have been at least partially vaccinated, and the percentage is considerably higher among the most vulnerable.[28] However, even with the level of vaccinations, COVID-19 continues to spread at least in part because of the much more contagious variants and in part because of geographic pockets of vaccine hesitancy. As of February 2022, Texas lags the national average in both partial and full vaccinations, as do many of the states in the South.[29] Collectively, this suggests that there will be

https://www.mayoclinic.org/coronavirus-covid-19/vaccine-tracker

[29] https://www.nytimes.com/interactive/2020/us/covid-19-vaccine-doses.html

significant pockets of communities where lack of restrictions – including the opening of schools – may still lead to considerable community spread moving forward.

For these reasons, debate about school openings and mitigation strategies will therefore likely continue to persist throughout the 2021-2022 school year, and our results provide important information that can help inform that debate. In particular, the CDC guidelines say that schools *can* reopen if community spread is low and considerable precautions are taken. Our study is not necessarily at odds with that guidance; instead, it simply shows that school reopenings *are not always* safe if those conditions are not met.

REFERENCES

Allcott, H., Boxell, L., Conway, J., Gentzkow, M., Thaler, M., & Yang, D. (2020). "Polarization and public health: Partisan differences in social distancing during the coronavirus pandemic." *Journal of Public Economics* 191:104254, doi:10.1016/j.jpubeco.2020.104254

Alsan, M. & Wanamaker, M. (2018). "Tuskegee and the Health of Black Men." *The Quarterly Journal of Economics* 133(1): 407–455, doi:10.1093/qje/qjx029

Andersen, M., Bryan, S., & Slusky, D. (2020). "COVID-19 Surgical Abortion Restriction Did Not Reduce Visits to Abortion Clinics." National Bureau of Economic Research Working Paper Series, No. 28058. doi:10.3386/w28058

Auger K.A., Shah S.S., Richardson T., et al. (2020) "Association Between Statewide School Closure and COVID-19 Incidence and Mortality in the US." *JAMA* 324(9):859–870. doi:10.1001/jama.2020.14348.

Bailey, M., Johnston, D.M., Koenen, M., Kuchler, T., Russel, D., & Stroebel, J. (2020). "Social Networks Shape Beliefs and Behavior: Evidence from Social Distancing During the COVID-19 Pandemic." National Bureau of Economic Research Working Paper Series, No. 28234. doi:10.3386/w28234

Baker, A.C., D. Larcker, and C.C.Y. Wang (2021). "How Much Should We Trust Staggered Difference-in-Differences Estimates." EGCI Finance Working Paper 736/2021.

Borusyak, Kirill and Jaravel, Xavier and Spiess, Jann, Revisiting Event Study Designs: Robust and Efficient Estimation (April 19, 2022). Available at SSRN: https://ssrn.com/abstract=2826228 or http://dx.doi.org/10.2139/ssrn.2826228

Bravata, D., J. H. Cantor, N. Sood, and C. M. Whaley (2021, April). Back to School: The Effect of School Visits During COVID-19 on COVID-19 Transmission. Working Paper 28645, National Bureau of Economic Research.

Bullinger, L.R., Carr, J.B., & Packham, A. (Forthcoming). "COVID-19 and Crime: Effects of Stay-at-Home Orders on Domestic Violence." *American Journal of Health Economics*, doi:10.1086/713787

Callaway, B. & Sant'Anna, P.H.C. (Forthcoming). "Difference-in-Differences With Multiple Time Periods." *Journal of Econometrics*. doi:10.1016/j.jeconom.2020.12.001

Cengiz, D., A. Dube, A. Lindner, and B. Zipperer (2019). "The Effect of Minimum Wages on Low-Wage Jobs." *The Quarterly Journal of Economics* 134 (3), 1405-1454.

Council of Economic Advisers. (2020). *Reopening Schools Is Key to Unlocking the Full Potential of America's Children*. The White House. https://www.whitehouse.gov/articles/reopening-schools-key-unlocking-full-potential-americaschildren/

Cotti, C., Engelhardt, B., Foster, J., Nesson, E., & Niekamp, P. (Forthcoming). "The relationship between in-person voting and COVID-19: Evidence from the Wisconsin primary." *Contemporary Economic Policy*, doi:10.1111/coep.12519

Courtemanche, C., Garuccio, J., Le, A., Pinkston, J., & Yelowitz, A. (2020). "Strong Social Distancing Measures in the United States Reduced the COVID-19 Growth Rate." *Health Affairs* 39(7): 1237–1246. doi:10.1377/hlthaff.2020.00608

Cronin, C.J., & Evans, W.N. (2020). "Private Precaution and Public Restrictions: What Drives Social Distancing and Industry Foot Traffic in the COVID-19 Era?" National Bureau of Economic Research Working Paper Series, No. 27531. doi:10.3386/w27531

Cunningham, S. (2021). *Causal Inference: The Mixtape*. New Haven, CT: Yale University Press.

Dave, D., Friedson, A. I., Matsuzawa, K., & Sabia, J. J. (2021). "When Do Shelter-In-Place Orders Fight COVID-19 Best? Policy Heterogeneity Across states and Adoption Time." *Economic Inquiry* 59(1): 29–52. doi:10.1111/ecin.12944

Dave, D.M., Friedson, A.I., Matsuzawa, K., Sabia, J.J., & Safford, S. (2020a). "Black Lives Matter Protests and Risk Avoidance: The Case of Civil Unrest During a Pandemic." National Bureau of Economic Research Working Paper Series, No. 27408. doi:10.3386/w27408

Dave, D., McNichols, D., & Sabia, J.J. (2020b). "The contagion externality of a superspreading event: The Sturgis Motorcycle Rally and COVID-19" *Southern Economic Journal* 87(3): 769–807. doi:10.1002/soej.12475

de Chaisemartin, C. & D'Haultfœuille, X. (2020). "Two-Way Fixed Effects Estimators with Heterogeneous Treatment Effects." *American Economic Review* 110(9): 2964–96. doi:10.1257/aer.20181169

Doyle T, Kendrick K, Troelstrup T, et al. (2021). "COVID-19 in Primary and Secondary School Settings During the First Semester of School Reopening — Florida, August–December 2020." MMWR Morb Mortal Wkly Rep 2021;70:437–441. doi:10.15585/mmwr.mm7012e2

Dub, T., Erra, E., Hagberg, L., Sarvikivi, E., Virta, C., Jarvinen, A., Osterlund, P., Ikonen, N., Haveri, A., Melin, M., Lukkarinen, T. J., & Nohynek, H. (2020). "Transmission of SARS-CoV-2 following exposure in school settings: Experience from two Helsinki area exposure incidents." *MedRxiv*, 2020.07.20.20156018. doi:10.1101/2020.07.20.20156018

Falk, A., Benda, A., Falk, P., Steffen, S., Wallace, Z., & Høeg, T.B. (2021) "COVID-19 Cases and Transmission in 17 K–12 Schools — Wood County, Wisconsin, August 31–November 29, 2020." MMWR Morb Mortal Wkly Rep 2021;70:136–140. doi:10.15585/mmwr.mm7004e3

Flaxman, S., Mishra, S., Gandy, A. et al. (2020). "Estimating the effects of non-pharmaceutical interventions on COVID-19 in Europe." *Nature* 584:257–261. doi:10.1038/s41586-020-2405-7

Fontanet, A., Grant, R., Tondeur, L., Madec, Y., Grzelak, L., Cailleau, I., Ungeheuer, M.N., Renaudat, C., Pellerin, S. F., Kuhmel, L., et al. (2020). "SARS-CoV-2 infection in primary schools in northern France: A retrospective cohort study in an area of high transmission." *MedRxiv*. doi:10.1101/2020.06.25.20140178

Friedson, A.I., McNichols, D., Sabia, J.J., & Dave, D. (Forthcoming). "Shelter-In-Place Orders and Public Health: Evidence From California During the COVID-19 Pandemic." *Journal of Policy Analysis and Management*, doi:10.1002/pam.22267

Gabriel T Bosslet, Micah Pollak, Jeong Hoon Jang, Rebekah Roll, Mark Sperling, Babar Khan, The Effect of In-Person Primary and Secondary School Instruction on County-Level

Severe Acute Respiratory Syndrome Coronavirus 2 Spread in Indiana, *Clinical Infectious Diseases*, Volume 74, Issue 1, 1 January 2022, Pages 17-23 https://doi.org/10.1093/cid/ciab306

Glaeser, E.L., Jin, G.Z., Leyden, B.T., & Luca, M. (2020). "Learning from Deregulation: The Asymmetric Impact of Lockdown and Reopening on Risky Behavior During COVID-19." National Bureau of Economic Research Working Paper Series, No. 27650. doi:10.3386/w27650

Goldhaber, D., Imberman, S., Strunk, K, Hopkins, B., Brown, N., Harbatkin, E., & Kilbride, T. (2021). "To What Extent Does In-Person Schooling Contribute to the Spread of COVID-19? Evidence from Michigan and Washington." American Institute for Research, working paper 247-2020-2. https://caldercenter.org/sites/default/files/CALDER%20WP%20247-1220-2.pdf

Goodman-Bacon, A. (Forthcoming). "Difference-in-Differences with Variation in Treatment Timing." *Journal of Econometrics*.

Green, D.A., Karimirad, A., Simard-Duplain, G., & Siu, H.E. (2020). "COVID and the Economic Importance of In-Person K-12 Schooling," National Bureau of Economic Research Working Paper Series, No. 28200. doi:10.3386/w28200

Gupta S, Georgiou A, Sen S, Simon K, Karaca-Mandic P. (2021). "US Trends in COVID-19–Associated Hospitalization and Mortality Rates Before and After Reopening Economies." *JAMA Health Forum.* 2(6):e211262. doi:10.1001/jamahealthforum.2021.1262

Gupta, S., Nguyen, T.D., Rojas, F.L., Raman, S., Lee, B., Bento, A., & Wing, C. (2020). "Tracking Public and Private Responses to the COVID-19 Epidemic: Evidence from State and Local Government Actions." National Bureau of Economic Research Working Paper Series, No. 27027. doi:10.3386/w27027

Harris, D., Ziedan, E., & Hassig, S. (2021). "The Effects of School Reopenings on COVID-19 Hospitalization." National Center for Research on Education Access and Choice. Available at: The-Effects-of-School-Reopenings-on-COVID-19-Hospitalizations-REACH-January-2021.pdf

Hsiang, S., Allen, D., Annan-Phan, S. et al. (2020). "The effect of large-scale anti-contagion policies on the COVID-19 pandemic." *Nature* 584, 262–267. doi:10.1038/s41586-020-2404-8

Honein, M.A., Barrios, L.C., & Brooks, J.T. (2021) "Data and Policy to Guide Opening Schools Safely to Limit the Spread of SARS-CoV-2 Infection." *JAMA.* 325(9):823–824. doi:10.1001/jama.2021.0374

Hurt, S., Ball, A., & Wedell, K. (2020). "Children more at risk for abuse and neglect amid coronavirus pandemic, experts say." USA Today, March 21, Accessed on April 14, 2021 at: https://www.usatoday.com/story/news/investigations/2020/03/21/coronavirus-pandemic-could-become-child-abuse-pandemic-experts-warn/2892923001/

Isphording, I.E., Lipfert, M., & Pestel, N. (2020). "School Re-Openings after Summer Breaks in Germany Did Not Increase SARS-CoV-2 Cases." IZA Institute of Labor Economics, IZA DP No. 13790. Retrieved from http://ftp.iza.org/dp13790.pdf

Kaufman, B. G., Mahendraratnam, N., Nguyen, T., Benzing, L., Beliveau, J., Silcox, C., & Wong, C.A. (2021). "Factors Associated with Initial Public School Reopening Plans During the US COVID-19 Pandemic: A Retrospective Study." Journal of General Internal Medicine, January. https://doi.org/10.1007/s11606- 020-06470-1.

Kniesner, T.J., & Sullivan, R. (2020) "The forgotten numbers: A closer look at COVID-19 non-fatal valuations." *Journal of Risk and Uncertainty* 61:155–176. doi:10.1007/s11166-020-09339-0

Kuhfeld, M., Soland, J., Tarasawa, B., Johnson, A., Ruzek, E., & Liu, J. (2020). "Projecting the potential impacts of COVID-19 school closures on academic achievement." (EdWorkingPaper: 20-226). Retrieved from Annenberg Institute at Brown University: https://doi.org/10.26300/cdrv-yw05

Kuhfeld, M., & Tarasawa, B. (2020). "The COVID-19 slide: What summer learning loss can tell us about the potential impact of school closures on student academic achievement." The Collaborative for Student Growth NWEA White Paper. https://www.nwea.org/research/publication/the-covid-19-slide-what-summer-learning-loss-can-tell-us-about-the-potential-impact-of-school-closures-on-student-academic-achievement/

Ludvigsson, J.F. (2020). "Children are unlikely to be the main drivers of the COVID-19 pandemic – A systematic review." *Acta Paediatrica* 109(8):1525–1530. doi:10.1111/apa.15371

Maldonado, J.E., & De Witte K. (2020) "The effect of school closure on standardized student test outcomes." Discussion paper series DPS20.17. Belgium: KU Leuven.

Milosh, M., Painter, M., Sonin, K., Van Dijcke, D., & Wright, A.L. (2020), "Unmasking Partisanship: How Polarization Influences Public Responses to Collective Risk", University of Chicago, Becker Friedman Institute for Economics Working Paper (2020-102)

MIT Election Data and Science Lab, (2018), "County Presidential Election Returns 2000-2016", https://doi.org/10.7910/DVN/VOQCHQ, Harvard Dataverse, V6, UNF:6:ZZe1xuZ5H2l4NUiSRcRf8Q== [fileUNF]

Poole, M.K., Fleischhacker, S., & Bleich, S.N. (2021). "Addressing Child Hunger When School Is Closed — Considerations during the Pandemic and Beyond." New England Journal of Medicine. https://doi.org/10.1056/NEJMp2033629.

Sass, T. and Thomas, G., (2021). "Student Achievement Growth During the COVID-19 Pandemic" *GPL Reports*. 9. https://scholarworks.gsu.edu/gpl_reports/9

Schmidt, S. & Natanson H. (2020). "With kids stuck at home, ER doctors see more severe cases of child abuse." *Washington Post*. Accessed on February 20, 2021 at: https://www.washingtonpost.com/education/2020/04/30/child-abuse-reports-coronavirus/

Simmons-Duffin, S.S. (2020). "14 States Make Contact Tracing Data Public. Here's What They're Learning." Available at https://www.npr.org/sections/health-shots/2020/08/14/902271822/13-states-make-contact-tracing-data-public-heres-what-they-re-learning

Simonov, A., Sacher, S. K., Dubé, J.P.H., & Biswas, S. (2020). "The Persuasive Effect of Fox News: Non-Compliance with Social Distancing During the Covid-19 Pandemic." National Bureau of Economic Research Working Paper Series, No. 27237. doi:10.3386/w27237

Valant, J. (2020). "School reopening plans linked to politics rather than public health." Brown Center Chalkboard, Brookings Institution. Accessed at https://www.brookings.edu/blog/brown-center-chalkboard/2020/07/29/school-reopening-plans-linked-to-politics-rather-than-public-health/

Table 1: Means and Standard Deviations of Outcome Variables

COVID Outcomes	(1) Pre-reopening	(2) Post-reopening
New cases per 100,000 residents	147.04	131.85
	(120.30)	(166.98)
New deaths per 100,000 residents	3.41	1.72
	(5.43)	(3.84)
Observations	2,024	2,277

Mobility Outcomes	Pre-reopening	Post-reopening
Time Completely Home (%)	28.21	26.62
	(5.94)	(5.72)
Part-time Work (%)	7.84	9.00
	(2.25)	(2.65)
Full-time Work (%)	3.98	4.93
	(1.35)	(1.69)
Median non-home dwelling time (minutes)	107.76	128.25
	(56.42)	(62.10)
Observations	87,480	102,060

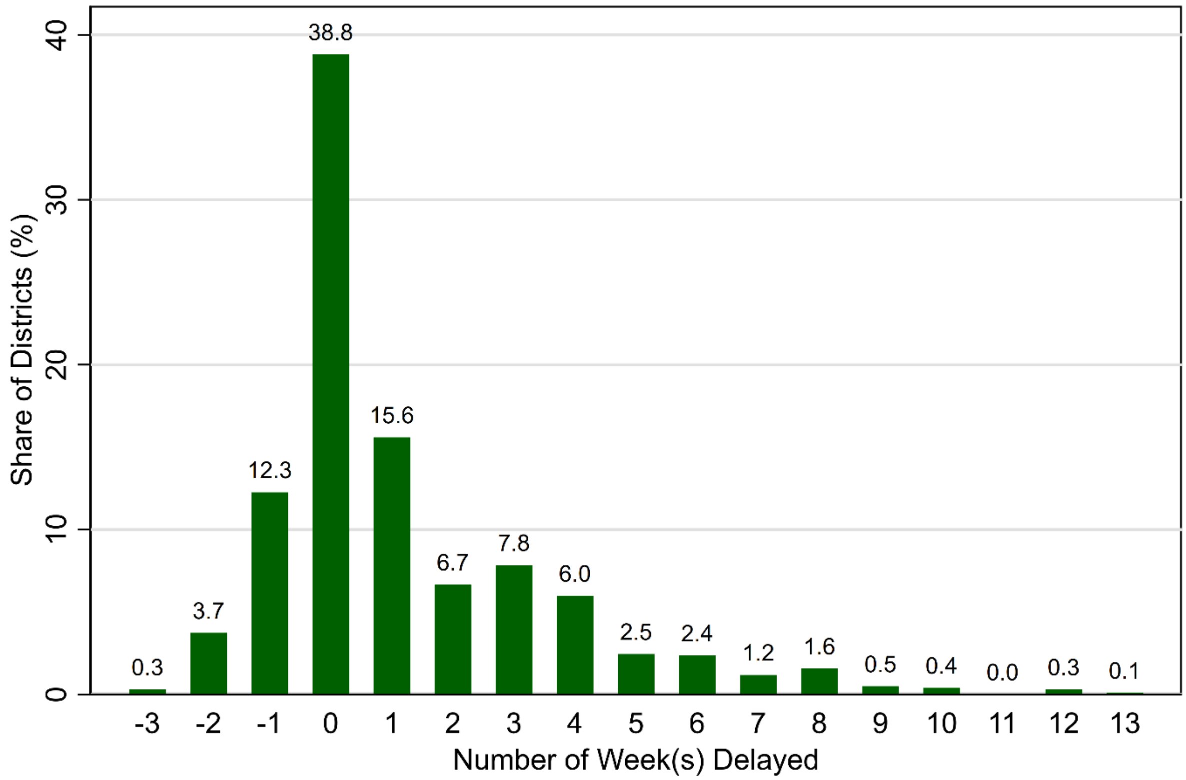
Notes: Standard deviations are in parentheses. The COVID outcomes utilize public county-by-week-level data, while the mobility outcomes are from census-block-group-by-week-level data from SafeGraph. Observations are weighted by county (census-block-group) population for the COVID (mobility) variables. The pre-reopening period refers to the eight (six) weeks prior to school reopenings for the COVID (mobility) variables. The post-reopening period refers to the reopening week along with the eight (six) weeks following reopening for the COVID (mobility) variables.

Table 2: Predictors of Reopening Week

	Coefficient estimate (standard error)
Standardized 2016 percent of votes for President Trump	-1.20***
	(0.30)
Standardized percent Hispanic	-0.01
	(0.15)
Standardized percent Black	-0.12
	(0.12)
Standardized population	0.19**
	(0.09)
Standardized percent who stayed at home for full day	0.01
	(0.15)
Standardized new weekly cases per 100,000	-0.11
	(0.48)
Constant	17.07***
	(0.33)
Observations	253

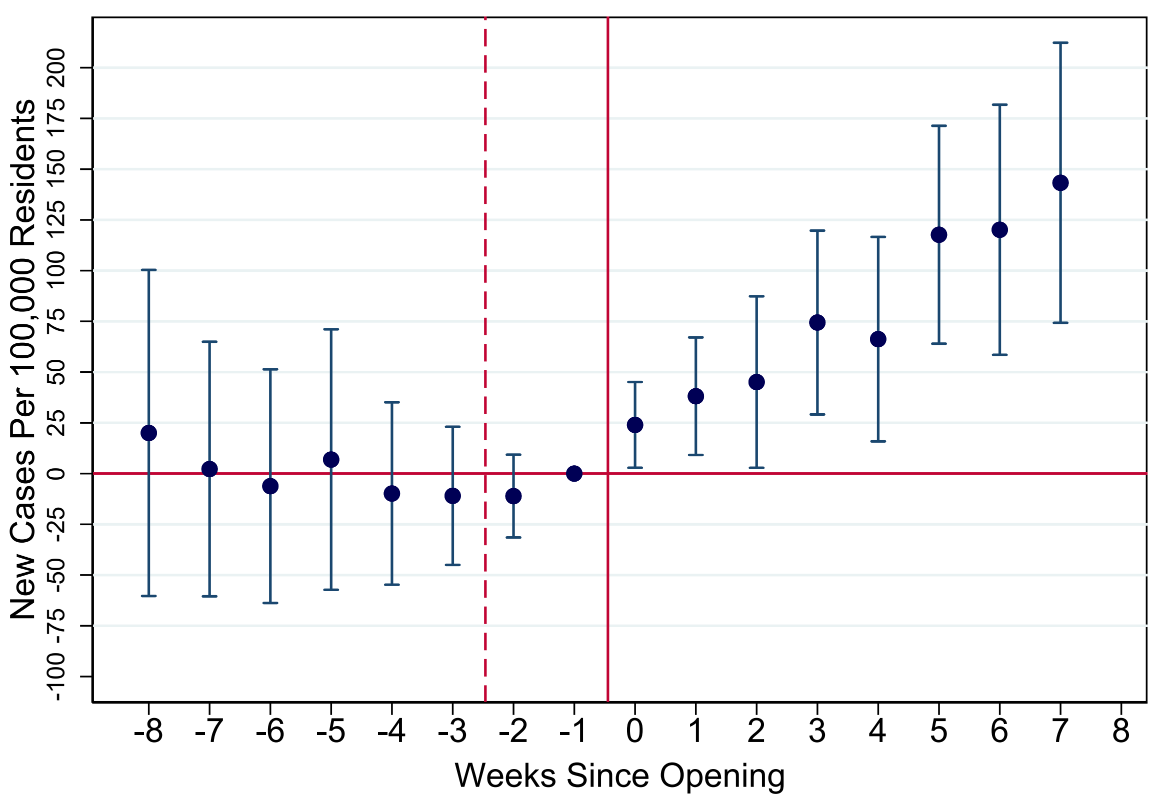
Notes: * p<0.1, ** p<0.05, *** p<0.01. Results are from a cross-sectional county-level linear regression with week number of reopening (ranging from 14 to 28, with 1 indicating the week of May 3) as the outcome variable. The stay-at-home and new cases variables are pooled averages across the four weeks prior to the earliest school reopening (week numbers 10 through 13).

Figure 1: Relative Start Date of School District Start Date in 2020-21 School Year Relative to the 2019-20 School Year



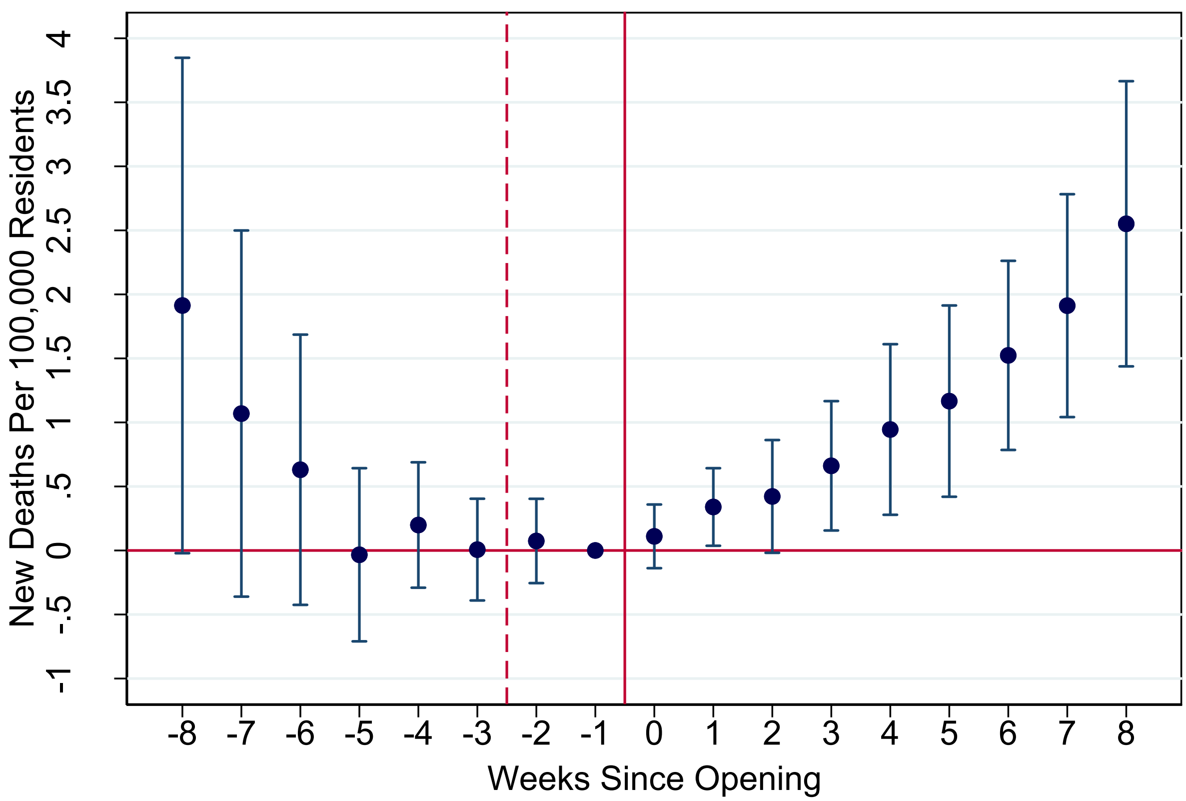
Note: In some cases, we do not have the 2019-20 start date for school districts. In these cases, we substitute a prior start date for any year we could find a record.

Figure 2: Event-Study Regression Results for Effect of Reopening Schools on COVID-19 Cases



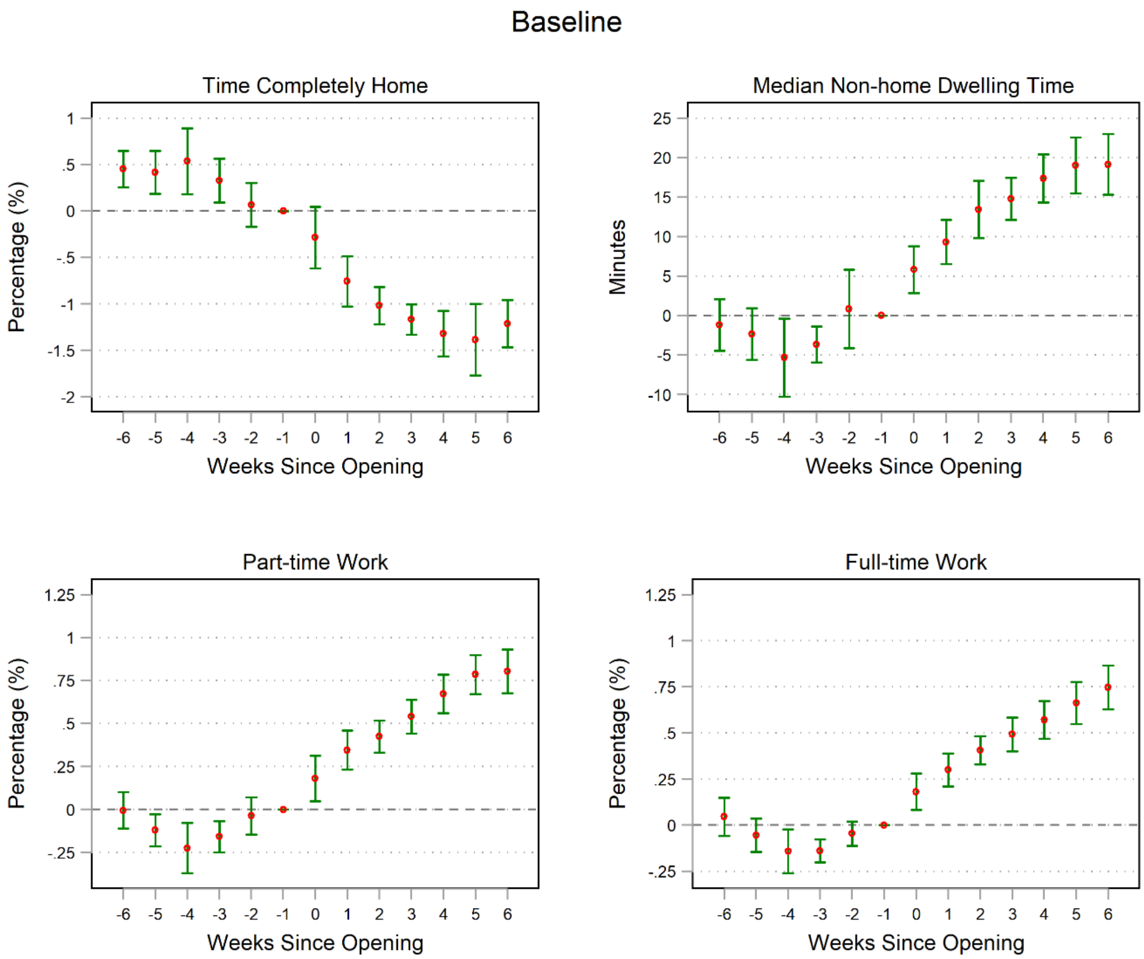
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure 3: Event-Study Regression Results for Effect of Reopening Schools on COVID-19 Fatalities per 100,000 Residents



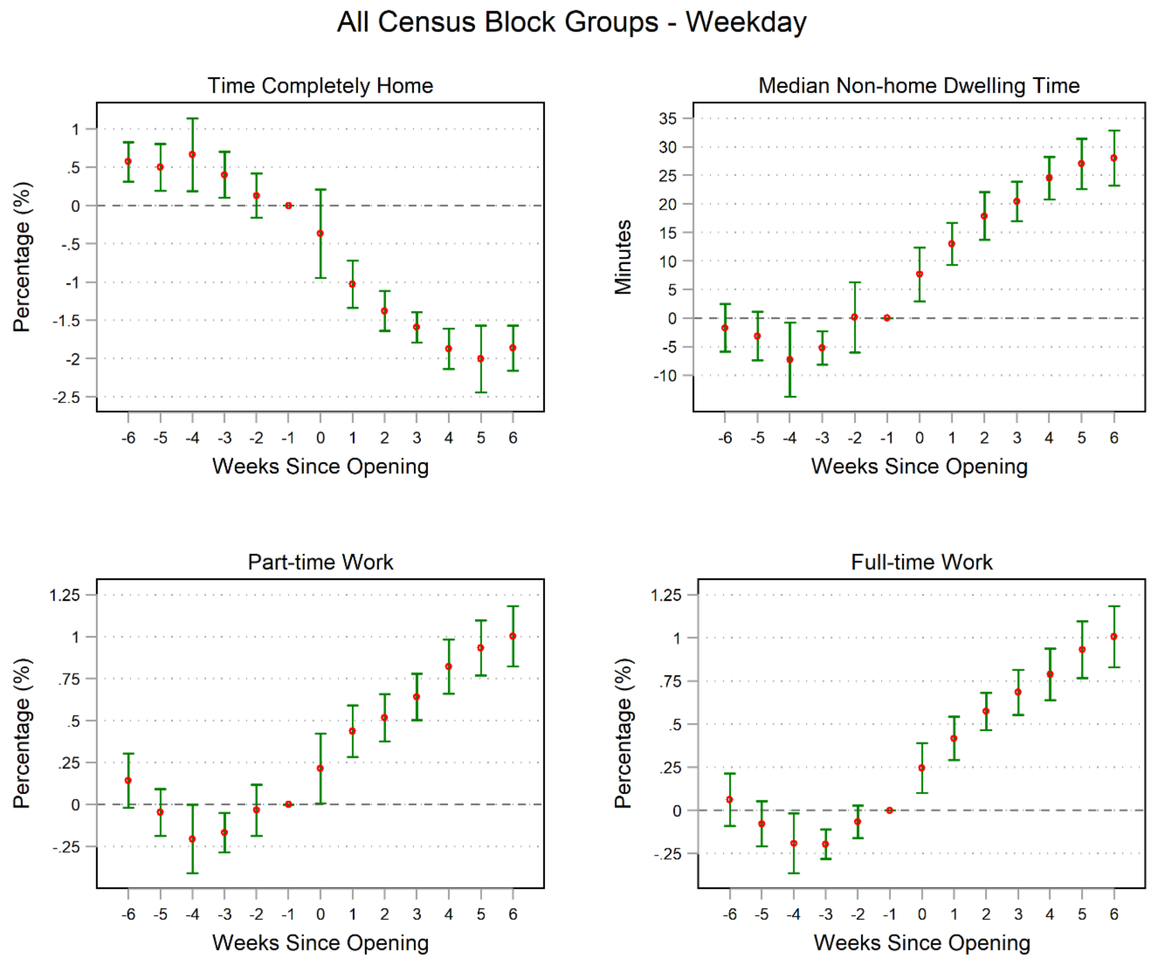
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure 4: Effects of School Reopening on Mobility - Full Week



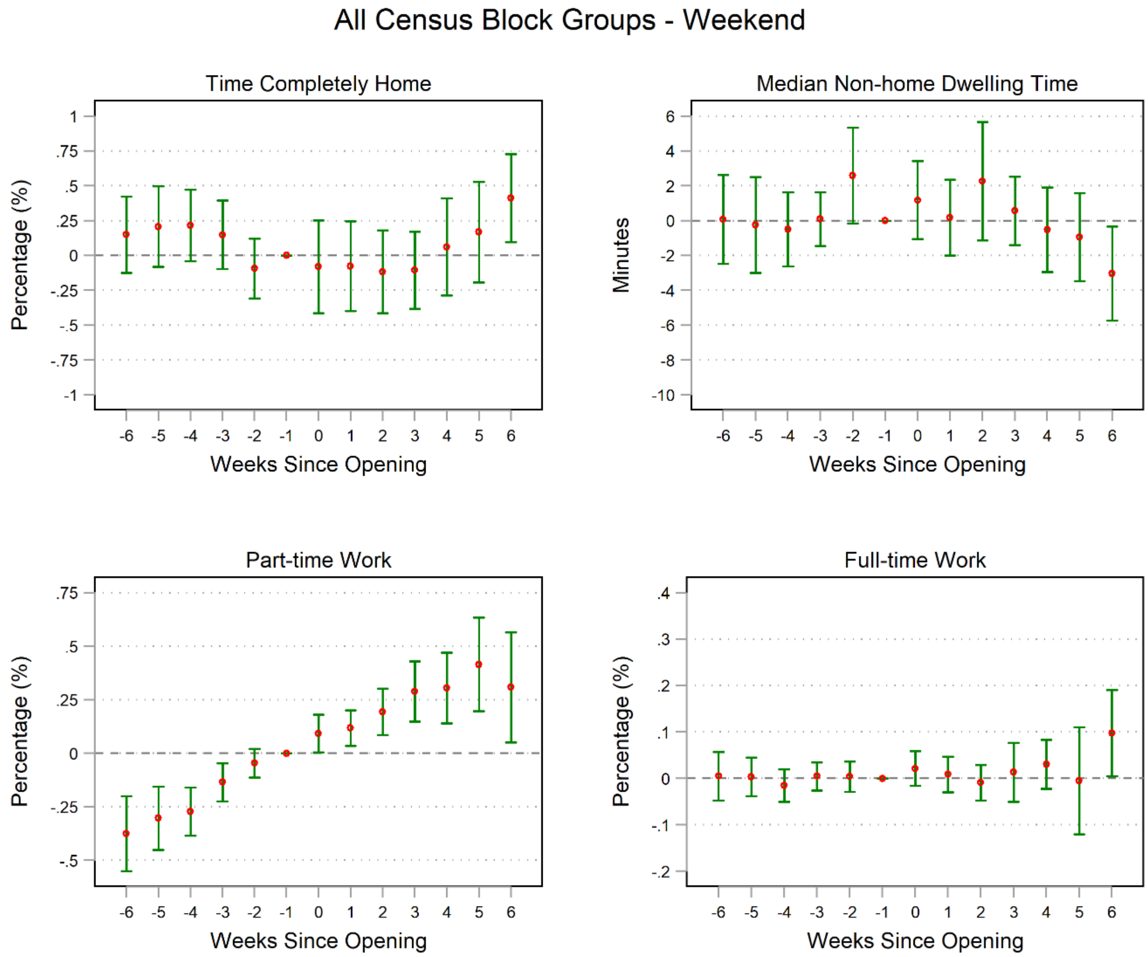
Notes: Results are from an event-study regression with SafeGraph census-block-group-by-week-level data from Texas. Dependent variable is at the top of each figure. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for fixed effects for census block group and week. Standard errors are heteroskedasticity-robust and clustered by county. N=247,860, or 14,580 census block groups x 17 weeks. Observations are weighted by census block group population.

Figure 5: Effects of School Reopening on Mobility - Weekday



Notes: Results are from an event-study regression with SafeGraph census-block-group-by-week-level data from Texas. Dependent variable is at the top of each figure. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for fixed effects for census block group and week. Standard errors are heteroskedasticity-robust and clustered by county. N=247,860, or 14,580 census block groups x 17 weeks (with weekends excluded). Observations are weighted by census block group population.

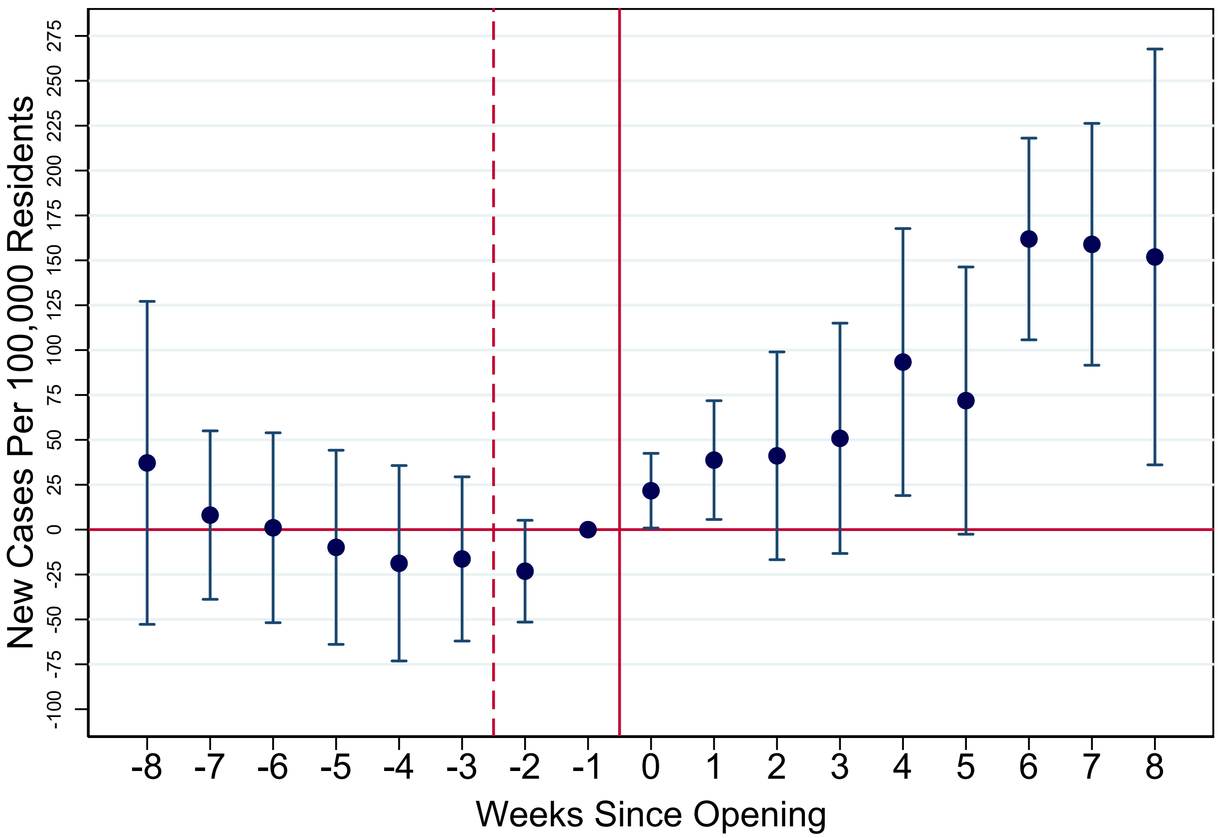
Figure 6: Effects of School Reopening on Mobility - Weekday



Notes: Results are from an event-study regression with SafeGraph census-block-group-by-week-level data from Texas. Dependent variable is at the top of each figure. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for fixed effects for census block group and week. Standard errors are heteroskedasticity-robust and clustered by county. N=247,860, or 14,580 census block groups x 17 weeks (with weekdays excluded). Observations are weighted by census block group population.

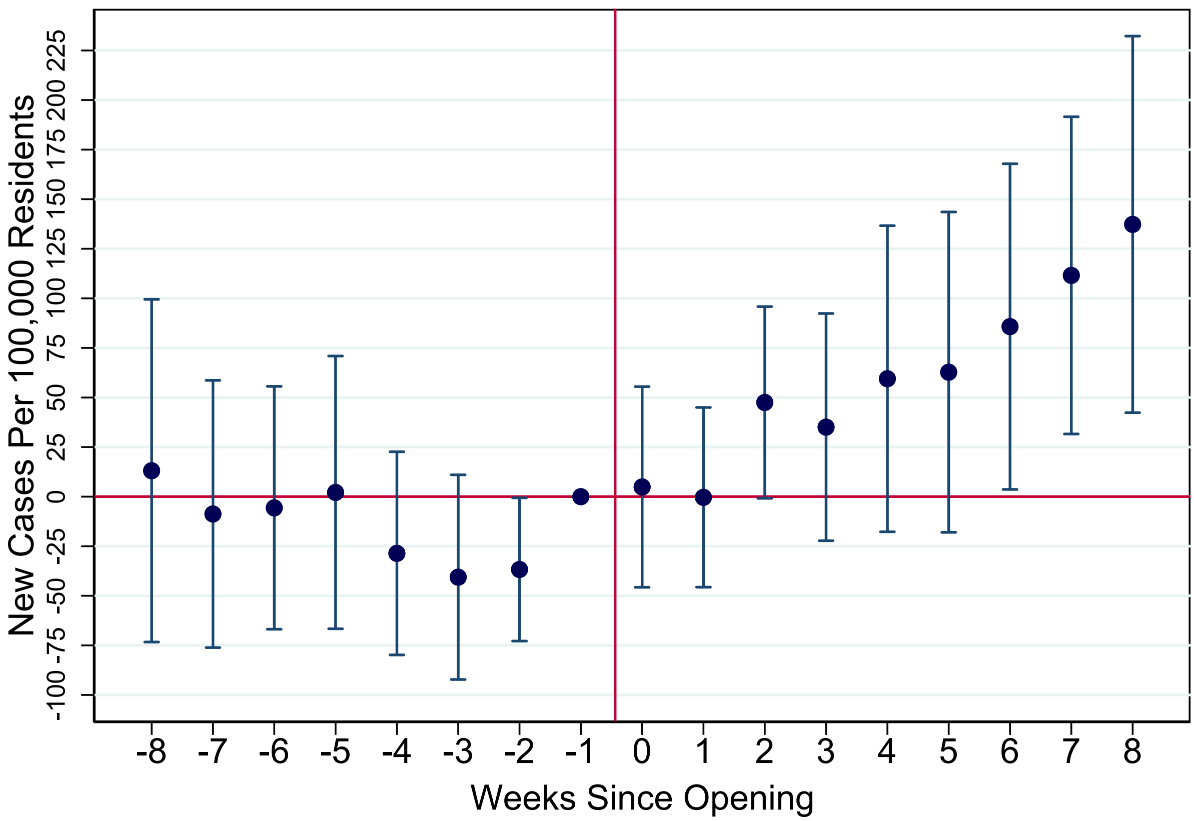
Appendix

Figure A1: Effect of Reopening Schools on COVID-19 Cases, Drop Counties with Substantial (40%+) Reopening Activity Prior to Treatment Week



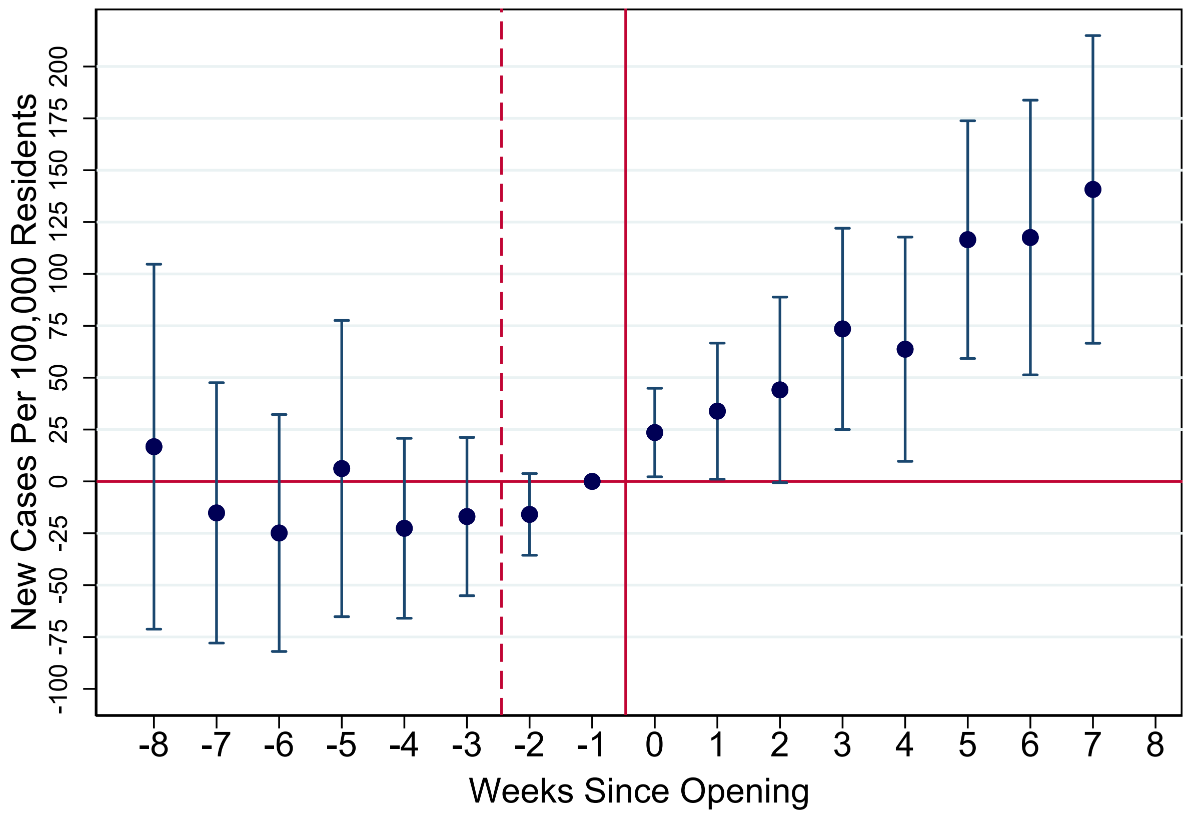
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.1 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. We exclude the 57 counties in which schools were already 40% reopened prior to what we classify as the treatment week. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=3,332, or 196 counties x 17 weeks. Observations are weighted by county population.

Figure A2: Effect of Reopening Schools on COVID-19 Cases, Treatment Defined as when the First School District Reopens in a County



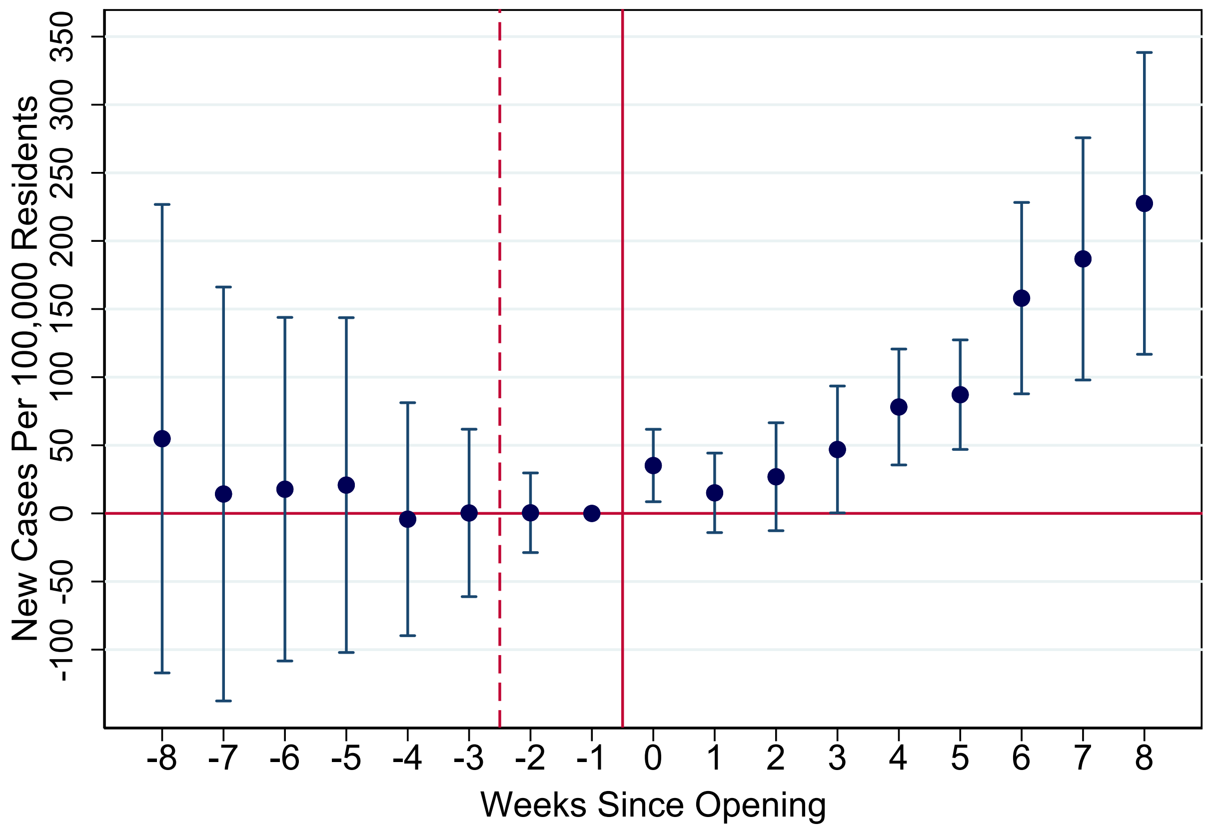
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when the first school district reopened for in-person learning in the county. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A3: Effect of Reopening Schools on COVID-19 Cases, Control for College Reopenings



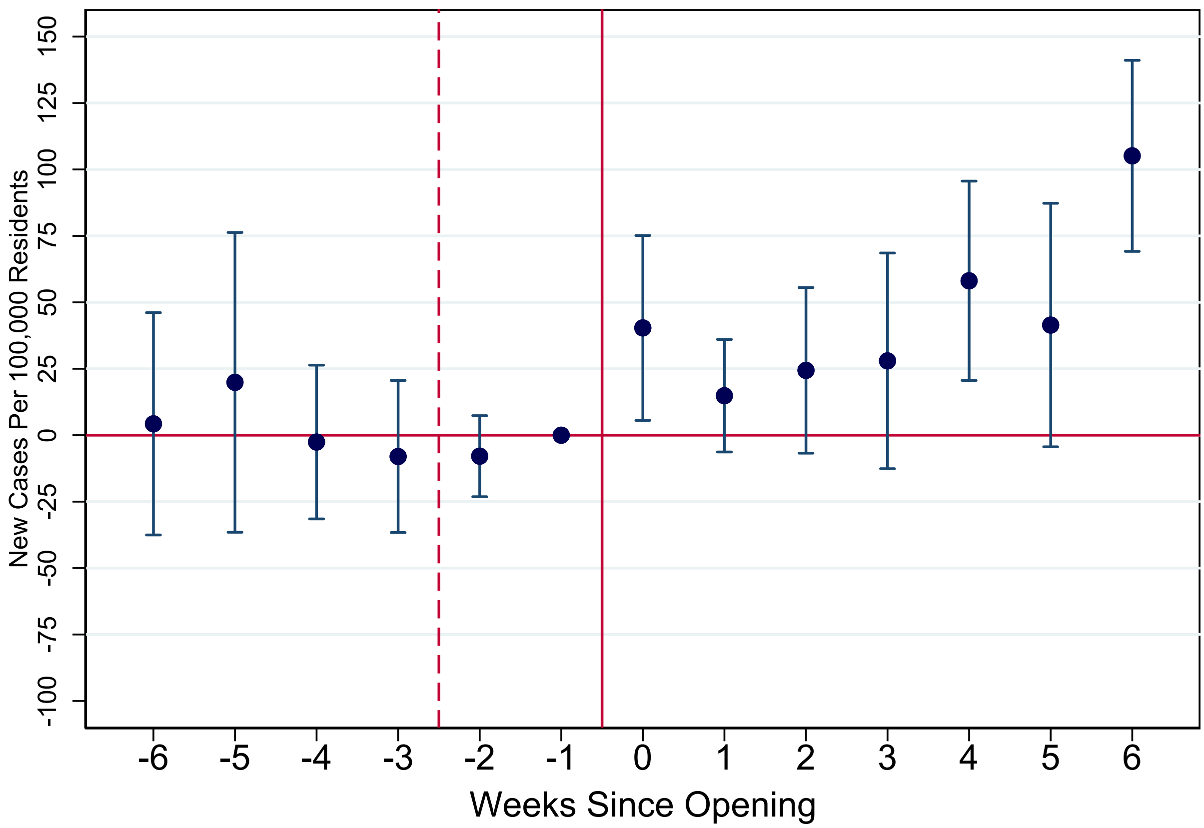
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for the proportion of the county's population comprised of college students interacted with an indicator for college being in session, new COVID-19 tests per 100,000 residents (as well as two of its lags), and fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A4: Effect of Reopening Schools on COVID-19 Cases, Control for Trump-Vote-Share-Specific Trends



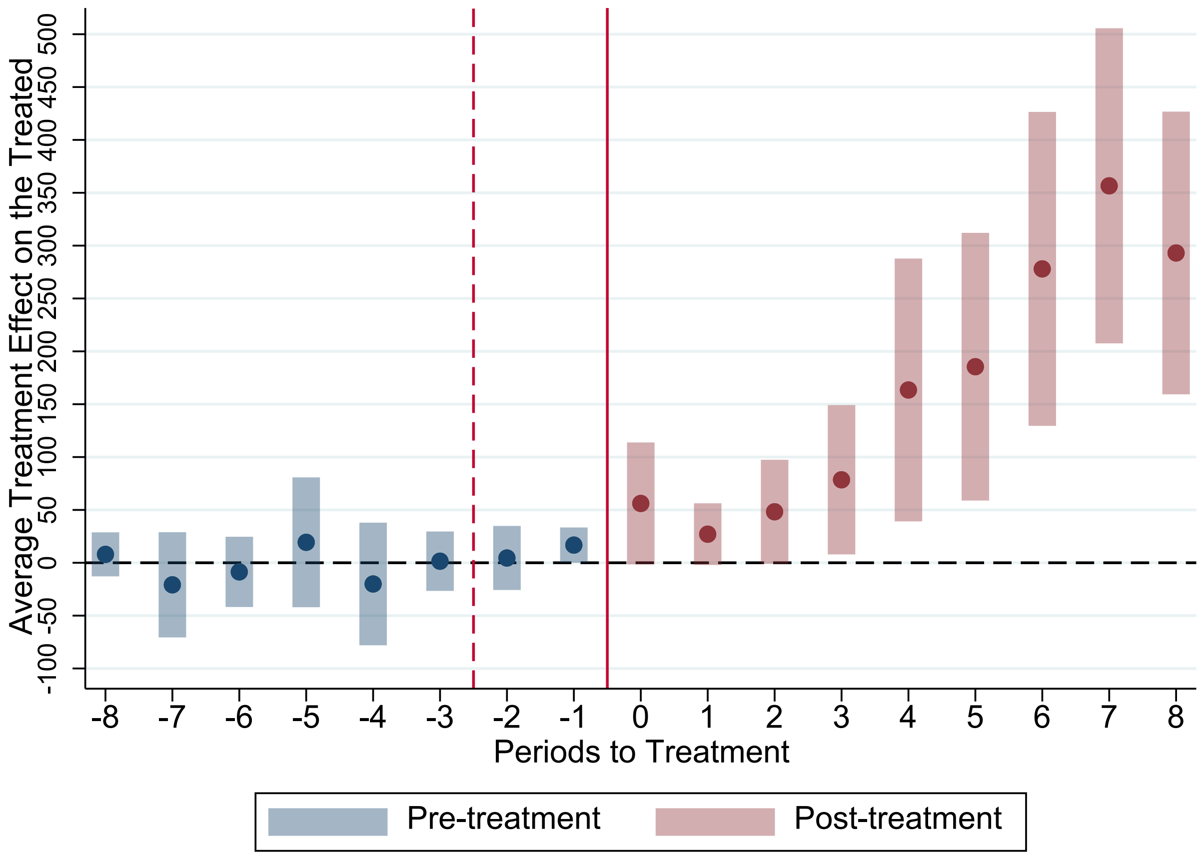
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for the interaction of a set of indicators reflecting quartile of Trump vote share with week fixed effects, new COVID-19 tests per 100,000 residents (as well as two of its lags), and fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A5: Effect of Reopening Schools on COVID-19 Cases, Six Week Window on Both Sides of Treatment



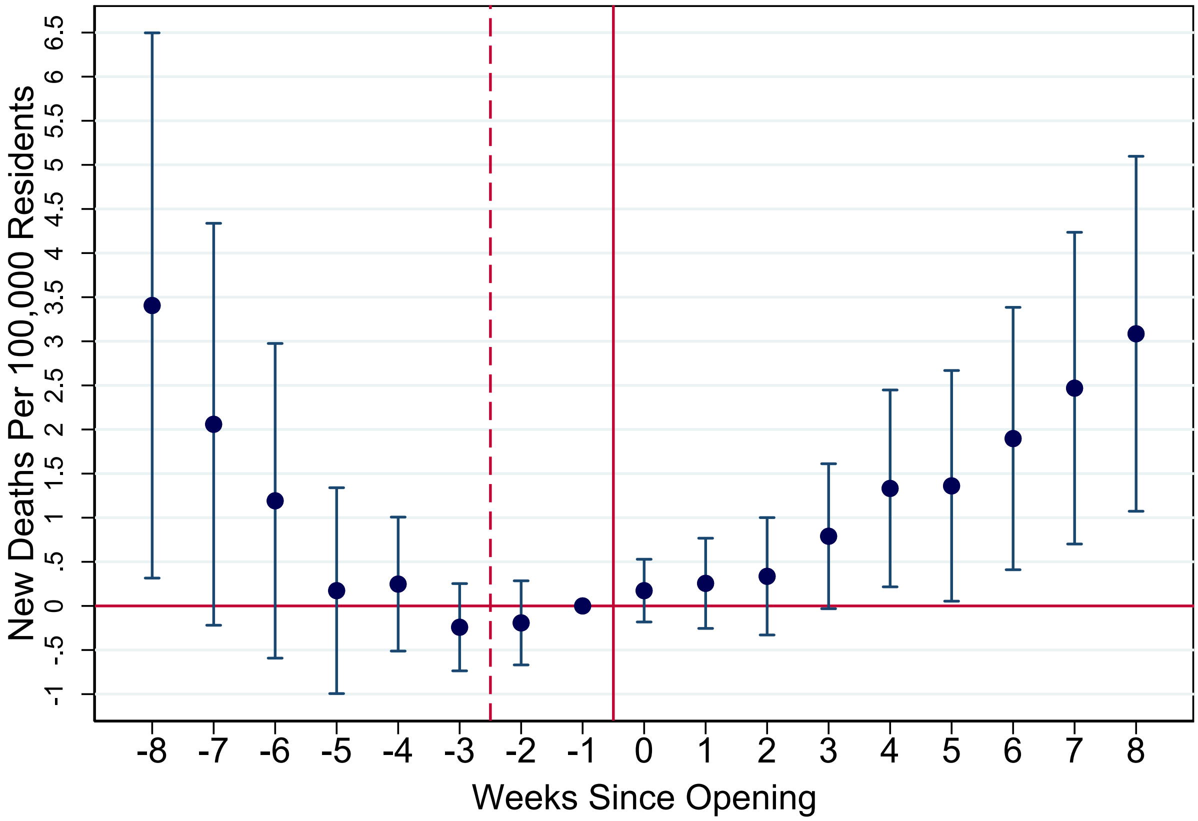
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for six weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=3,289, or 253 counties x 13 weeks. Observations are weighted by county population.

Figure A6: Effect of Reopening Schools on COVID-19 Cases per 100,000 Residents, Callaway-Sant'Anna Approach



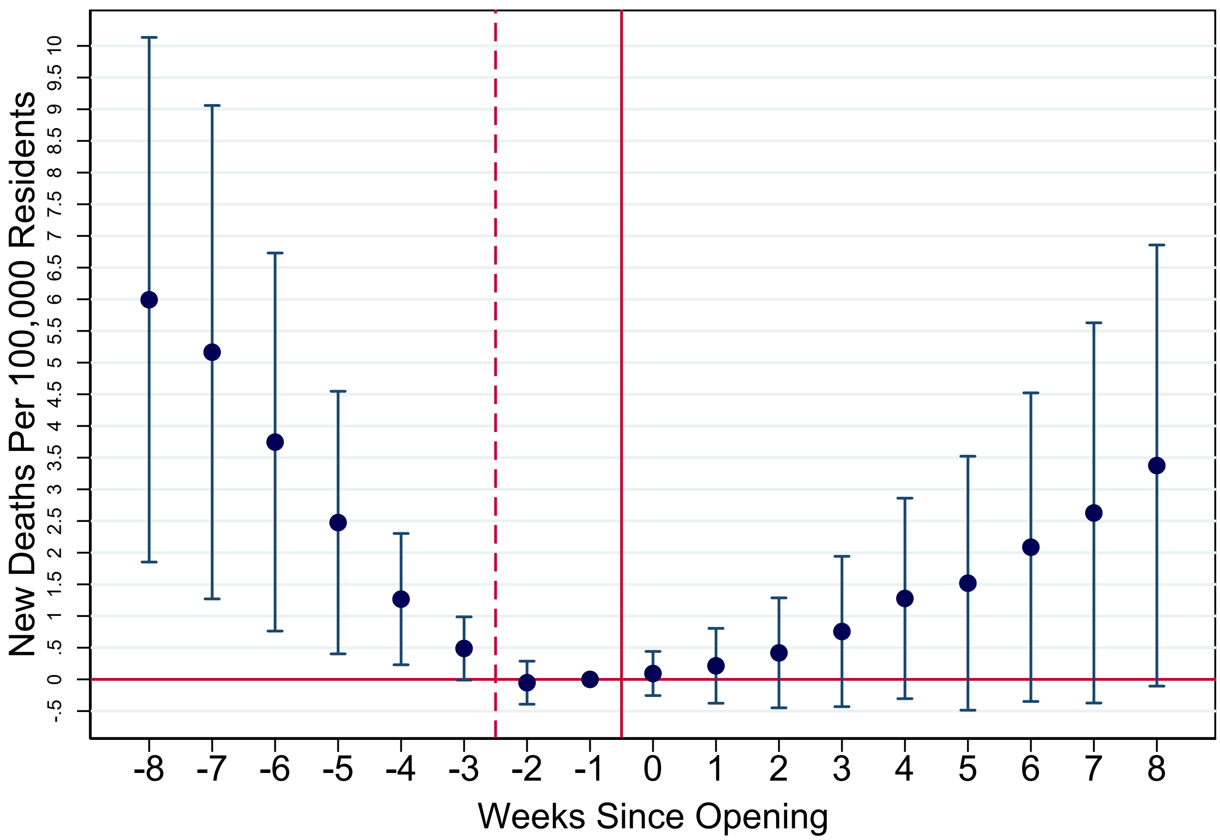
Notes: Results are from a Callaway-Sant'Anna-style event-study regression with county-by-week-level data from Texas. New COVID-19 cases per 100,000 residents is the dependent variable. In the pre-treatment period, estimates shown, along with 95% confidence intervals, are for the change in treatment effect from the previous period (e.g. for -8 this represents the change from -9 to -8). This is why there is no reference period. In the post-treatment period, estimates shown are indicators for each of eight weeks since treatment. Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for fixed effects for county and week. Standard errors are bootstrapped. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A7: Effect of Reopening Schools on COVID-19 Fatalities, Drop Counties whose Largest Increase in Reopening Occurred Earlier than when Our Method Defines Treatment



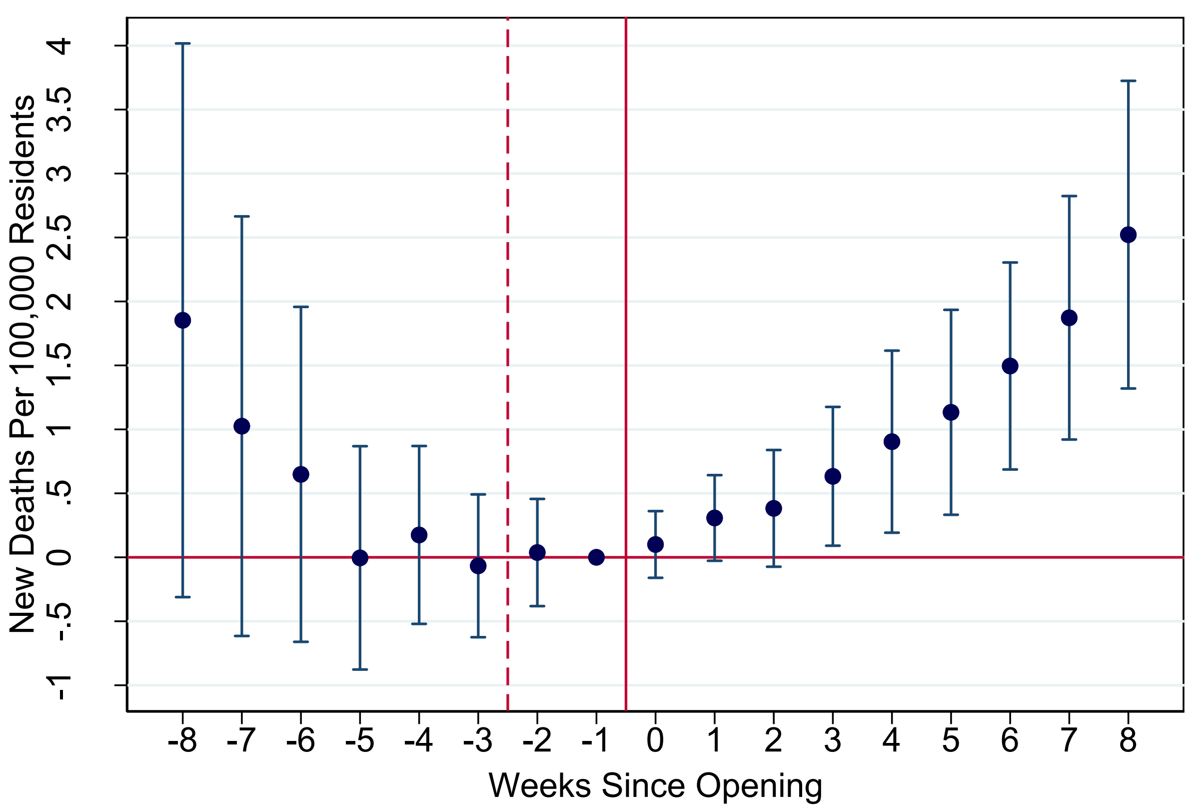
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. We exclude the 57 counties in which schools were already 40% reopened prior to what we classify as the treatment week. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=3,332, or 196 counties x 17 weeks. Observations are weighted by county population.

Figure A8: Effect of Reopening Schools on COVID-19 Fatalities, Treatment Defined as when the First School District Reopens in a County



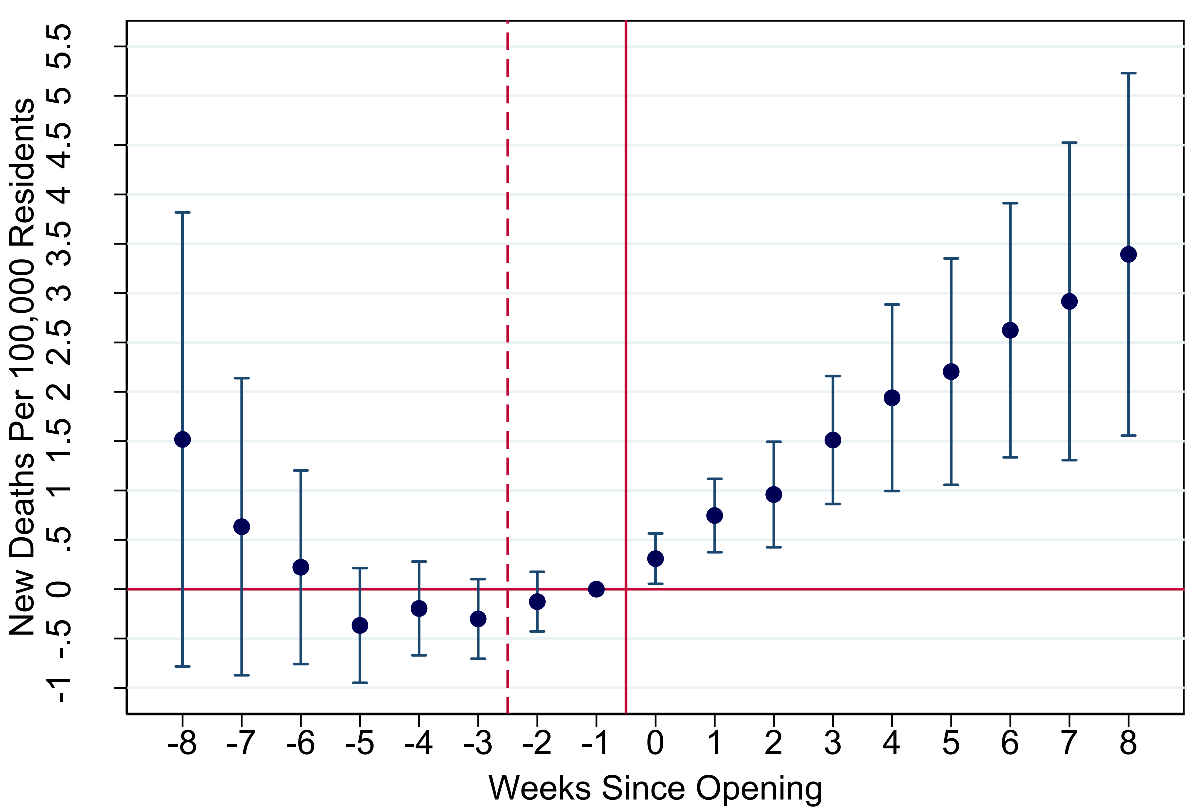
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A9: Effect of Reopening Schools on COVID-19 Fatalities, Control for College Reopenings



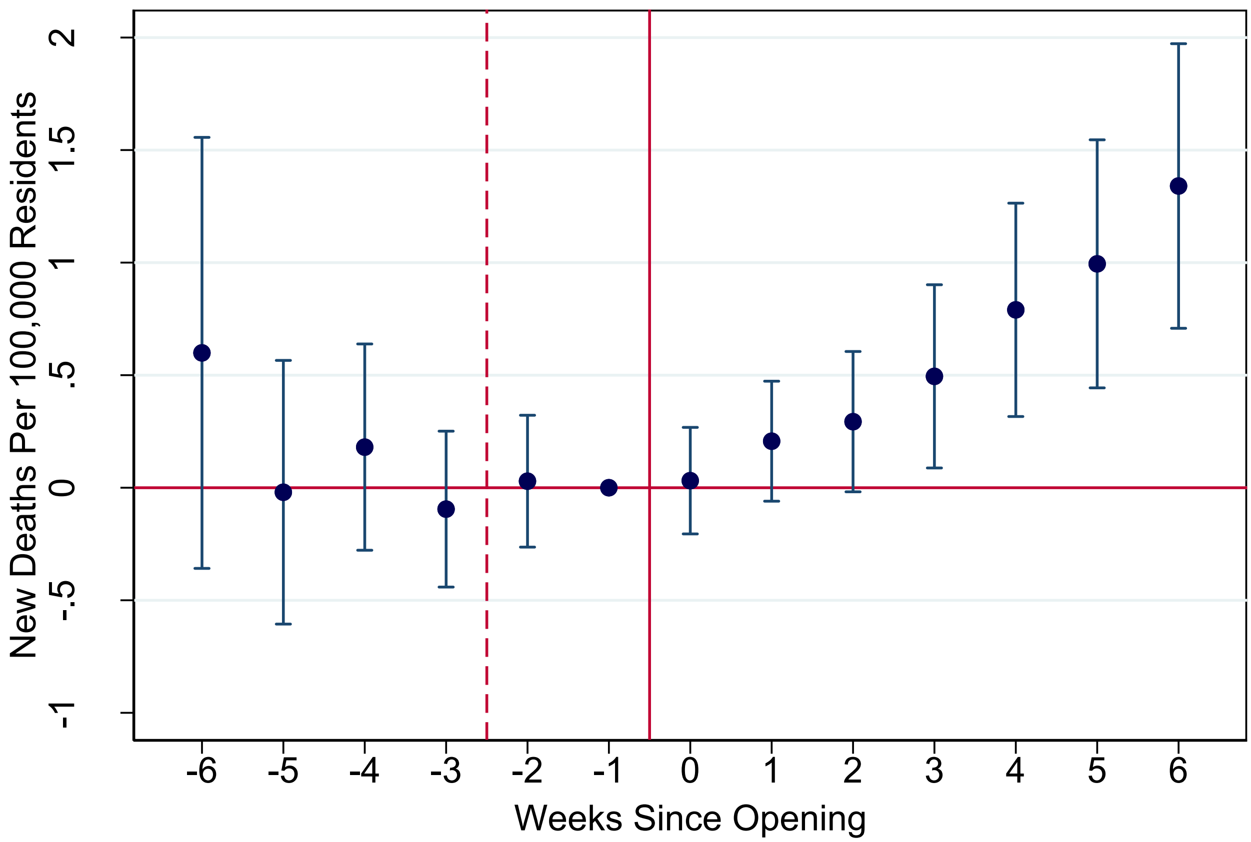
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for the proportion of the county's population comprised of college students interacted with an indicator for college being in session, new COVID-19 tests per 100,000 residents (as well as two of its lags), and fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A10: Effect of Reopening Schools on COVID-19 Fatalities, Control for Trump-Vote-Share-Specific Trends



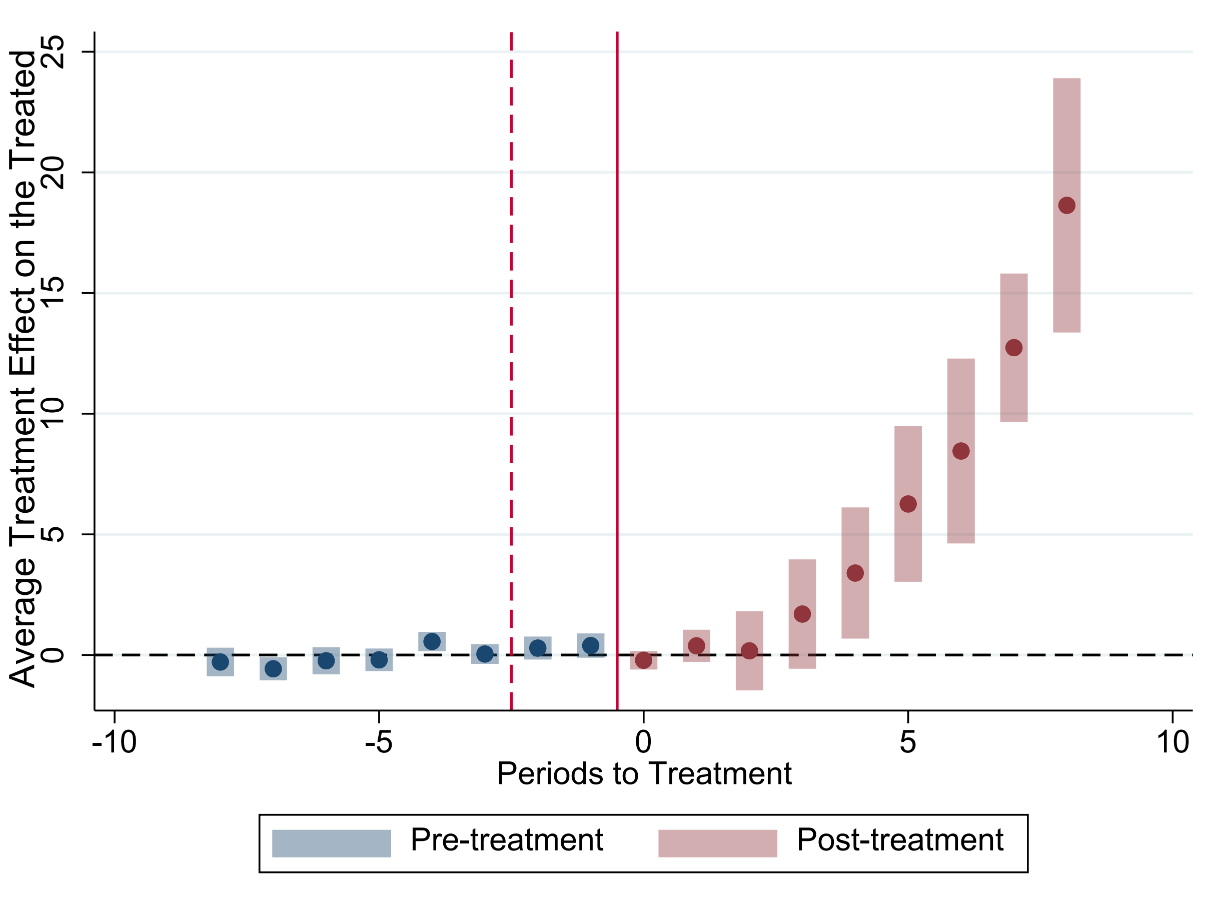
Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for eight weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for the interaction of a set of indicators reflecting quartile of Trump vote share with week fixed effects, new COVID-19 tests per 100,000 residents (as well as two of its lags), and fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.

Figure A11: Event-Study Regression Results for Effect of Reopening Schools on COVID-19 Fatalities, Six Week Window on Both Sides of Treatment



Notes: Results are from an event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. Estimates shown, along with 95% confidence intervals, are for indicators for six weeks on each side of treatment (with the week before treatment being the reference period). Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for new COVID-19 tests per 100,000 residents (as well as two of its lags) along with fixed effects for county and week. Standard errors are heteroskedasticity-robust and clustered by county. N=3,289, or 253 counties x 13 weeks. Observations are weighted by county population.

Figure A12: Event-Study Regression Results for Effect of Reopening Schools on COVID-19 Fatalities, Callaway-Sant'Anna Approach



Notes: Results are from a Callaway-Sant'Anna-style event-study regression with county-by-week-level data from Texas. New COVID-19 deaths per 100,000 residents is the dependent variable. In the pre-treatment period, estimates shown, along with 95% confidence intervals, are for the change in treatment effect from the previous period (e.g. for -8 this represents the change from -9 to -8). This is why there is no reference period. In the post-treatment period, estimates shown are indicators for each of eight weeks since treatment. Treatment, denoted by week=0 and the solid vertical line, is defined as the week when more than 50% of students attended schools that were reopened for in-person learning. The dashed vertical line represents the fact that the average county's (weighted by population) first district reopened 2.5 weeks prior to the week it crossed the 50% threshold; therefore, "early" treatment effects are plausible. The regression controls for fixed effects for county and week. Standard errors are bootstrapped. N=4,301, or 253 counties x 17 weeks. Observations are weighted by county population.