Reply to the Comments on "The Association of Opening K-12 Schools and Colleges with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis"

Reply to Editor

Thank you very much for your helpful and constructive comments on our paper. We appreciate the time you spent in reviewing this paper. We have followed your recommendations, and those of the other reviewers', as far as possible. Major changes are:

1. Following the suggestion by the editor and the referees, we added the event-study regression analysis. Fig. 2 and 3 indicate that the gap in the average level of cases/deaths between in-person/hybrid school opening and remote school opening grows larger as more weeks pass after the school openings. We also provide the robustness checks in the panel regression analysis by estimating a specification with the log of weekly cases in place of the log difference as outcome variables.

2. Following the advice of the editor and the referee, we rearrange the order of our presentation by first discussing the data and the definition of variables in details and then providing a summary table with descriptive analysis before conducting the event-study analysis and the panel regression analysis.

3. We explicitly write a regression model which is motivated by the SIRD model, and we clearly state the unconfoundedness assumption under which we may interpret our panel regression result as causal.

4. Given the comment of the referee that "the evidence regarding colleges is substantially less strong than the evidence for K-12 schools" as well as the space constraint after adding the event-study analysis, we decided to drop our focus on the effect of college openings and mainly focus on the effect of K-12 school openings in our revised manuscript. In particular, we made the following changes.

   - We moved Figure 2 in the previous manuscript and its related discussion on the effect of college openings on cases and deaths to SI Appendix.
   - We moved Table 2(b) in the previous manuscript on the effect of opening colleges and K-12 schools on the visits to restaurants and bars to SI Appendix because Table 2(b) was presented to highlight the effect of college openings on an increase in the visits to bars.
   - The paper's title has been changed to "The Association of Opening K-12 Schools with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis" by dropping "and Colleges" from the previous title.

5. For the death growth regression, we now use the outcome variable defined by the log difference over 21 days in weekly deaths rather than the log difference over 7 days because the available evidence from the CDC shows that the time lag between infection and death reporting is stochastic and spreads over at least 2 weeks (c.f., Table 2 of `https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html`). We also use the explanatory variables that are lagged by 35 days to capture the time lag of infection and death reporting, where we provide robustness checks with respect to the choice of lags in sensitivity analysis.

6. We now use the standard fixed effects estimator without bias correction rather than debiased fixed effects estimator for the effect of school openings on full-time workplace visits, staying home devices, visits to restaurants and bars in Tables 3 and Table S4 in the revised manuscript (which corresponds to Table 2(a)(b) in the previous manuscript). This is because the specification for mobility outcome does not have lagged dependent variables in the covariates and there is no need for implementing bias-corrections. The results are similar after changing from the debiased fixed effects estimator to the standard fixed effects estimator.

Reply to Editor's Comments

- *"The comment that comes out clearly in both reports is that the assumptions behind the reasoning that your effects are causal is not elucidated. Writing down a model with the required interrelationship between the variables, including the disturbance terms, would go a long way towards answering a number of issues specified in the reports. It would also allow you to address many of the issues raised by ref no. 2. No model is perfect, but we need something to help us understand what is required to rely on your assumptions."*

  Thank you for your suggestion on presenting a model with the required interrelationship between the variables.

  Following your suggestion, we derive a regression specification from the SIRD model and present it in the main text as follows:

$$\Delta_7 \log Case_{it} = \beta' \, Visit_{i,t-14} + \sum_{\tau=14,21,28} \beta_{y,\tau} \log Case_{i,t-\tau} + \gamma' NPI_{i,t-14} + \theta \, Test_{it} + \alpha_i + \delta_{s(i),w(t)} + \epsilon_{it}, \quad [5]$$

  where i is county, t is day. The outcome variable $\Delta_7 \log Case_{it} := \log Case_{it} - \log Case_{i,t-7}$ is the log-difference over 7 days in reported weekly cases with $Case_{it}$ denoting the number of confirmed cases from day $t-6$ to t. Here, α_i and $\delta_{s(i),w(t)}$ represent county fixed effects and state-week fixed effects.

Furthermore, we now discuss the assumption that is required for interpreting our estimate as the causal effect as follows:

> For parameter identification, we assume that the error ϵ_{it} in [5] is orthogonal to the observed explanatory variables of school visits/openings, NPIs, test rates, and the past log cases, county fixed effects, and state-week fixed effects. The estimated parameters for school openings can be causally interpreted under the unconfoundedness assumption that the variables related to school openings (school visits, opening dates, teaching methods) are as good as randomly assigned after conditioning on other controls, county fixed effects, and state-week fixed effects.

The key assumption for our panel data analysis is that the timing and the teaching mode of school opening is random conditional on our specified confounders (county fixed effects, state-week fixed effects, observed policy variables such as mask mandates, stay-at-home, and ban gathering, observed mobility variables including visits to restaurants, bars, gyms, churches with various lag length). This unconfoundedness assumption is different from the parallel trend assumption in the event-study or diff-in-diff design.

SI Appendix, The Model and Methods provide further details.

- *"There is also requests for some additional reduced form analysis. Of particular interest would be the level regression, perhaps in logs but also just in numbers (as this gets rid of the log zero problem). Your growth rate specification is essentially a differential equation which depends on levels. As a result I think some reporting of the implications for the time trends from various policies would be interesting."*

  Following the suggestion from you and the referees, we now present the result of event-study design using the number of new confirmed weekly cases per 1000, the log of weekly cases, the number of new confirmed weekly deaths per 1000, and the log of weekly deaths as outcome variables as shown in Figures R1 and R2. In in Figures R1, we divide the sample into different subsamples, where each subsample contains the observation with the similar school opening dates and then, for each of subsamples, we run the following event-study regression with weekly dummies of leads and lags for three school opening modes (i.e., in-person, hybrid, and remote) and county fixed effects but without time-fixed effects:

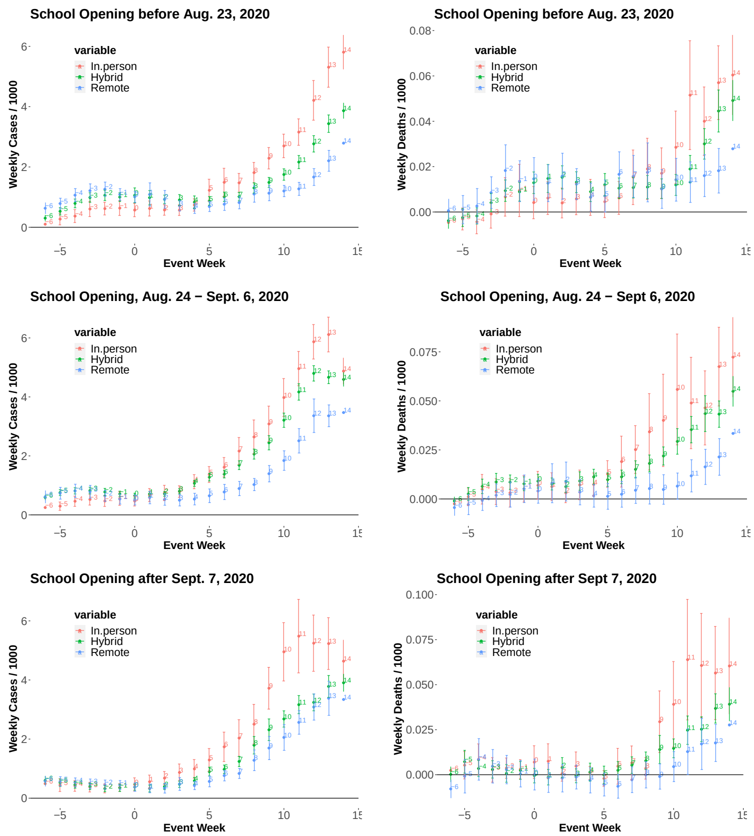
$$Y_{it} = \sum_{p \in \{\text{in-person,hybrid,remote}\}} \sum_{w=-8}^{22} \gamma_w^p D_{\tau,it}^p + \alpha_i + \epsilon_{it}, \tag{1}$$

where $D_{\tau,it}^p$ takes the value equal to 1 if school has been opened for τ weeks (or will be opened after $-\tau$ weeks if $\tau < 0$) with teaching mode p in county i at day t. In Figure

R4, we follow the proposed method of Callaway and Sant'Anna (2020) using their did R package to estimate the difference in the dynamic treatment effects between remote school openings and in-person/hybrid school openings. The result can be causally interpretable under the group-by-group parallel trend assumption.

Figure R1: The event-study regression estimates before and after the opening of K-12 schools

(a) Cases by Opening Modes (b) Deaths by Opening Modes

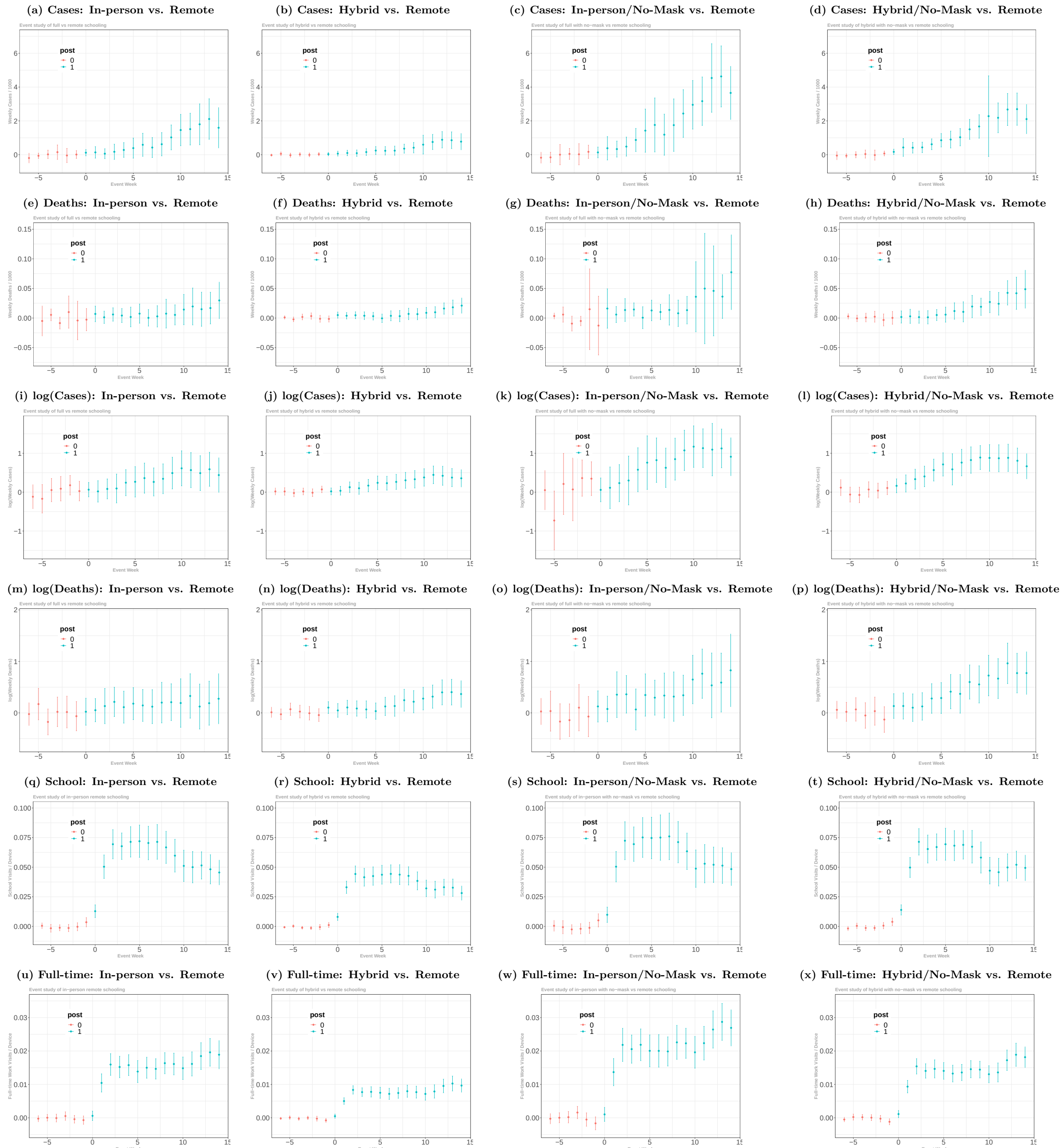


Notes: The figures plot the estimated coefficients for weekly dummies of leads and lags in the regression specification (1) with 95 percent confidence intervals for cases per 1000 and deaths per 1000 as outcome variables for three subsamples classified by school opening dates.

As shown in Figures R1 and R2, the result indicates that the gap in the average level of cases/deaths between in-person/hybrid school opening and remote school opening grows larger as more weeks passes after the school openings. See also our reply to Referee 1 for details. We would like to thank both the editor and the referees for encouraging us to present these results because we believe that these results reinforce our findings.

The pre-treatment estimates suggest no clear violation of the parallel trend assumption in pre-treatment period. On the other hand, our view on the necessity of parallel trend assumption for our dynamic panel regression analysis is different from one of the referees.

Figure R2: The average treatment estimates obtained using the DID method from Callway and Sant'Anna (2020)



Notes: (a) plots the estimates of the average dynamic treatment effect of in-person openings relative to the counties with remote openings as well as the counties that have not opened yet on cases per 1000 using a subset of counties with either in-person opening or remote-opening, where we use the estimation method of Callaway and Sant'Anna (2020) implemented by their did R package. Similarly, (b), (c), and (d) plots the estimates of the average dynamic treatment effect of school opening with hybrid, in-persion/mask mandates, and hybrid/mask mandates teaching methods, respectively, using a subset of counties with the corresponding teaching method as well as remote-opening. (e)-(h), (i)-(l), (m)-(p), (q)-(t), and (u)-(x) report the estimates of the average dynamic treatment effect on deaths per 1000, log(cases), log(deths), per-devise visits to K-12 schools, and per-devise visits to full-time workplaces, respectively.

Our panel regression design is a fundamentally different research design that requires the unconfoundedness assumption that the timing and the mode of school openings are random conditional on controlled confounders.

Table R1: Summary Statistics

	Wkly Case Growth	Wkly Death Growth	Wkly Cases per 1000	Wkly Deaths per 1000	K-12 Sch. Visits	Workplace Visits	Restaurant Visits
In-person							
Before Opening							
Mean	0.091	0.013	0.571	0.060	0.045	0.047	0.185
	(0.011)	(0.003)	(0.031)	(0.003)	(0.002)	(0.001)	(0.007)
N	52258	52258	54995	54995	67070	67070	67070
After Opening							
Mean	0.143	0.034	3.038	0.104	0.161	0.073	0.188
	(0.014)	(0.006)	(0.200)	(0.005)	(0.005)	(0.001)	(0.006)
N	45749	45749	45827	45827	46030	46030	46030
Difference in Means	0.052	0.021	2.467	0.044	0.116	0.026	0.003
	(0.018)	(0.007)	(0.203)	(0.004)	(0.004)	(0.001)	(0.004)
Hybrid							
Before Opening							
Mean	0.096	0.024	0.664	0.035	0.036	0.045	0.242
	(0.012)	(0.006)	(0.031)	(0.001)	(0.001)	(0.0004)	(0.005)
N	234820	234820	243321	243321	260573	260551	260573
After Opening							
Mean	0.121	0.042	2.368	0.057	0.126	0.064	0.249
	(0.012)	(0.006)	(0.132)	(0.002)	(0.003)	(0.001)	(0.004)
N	166605	166605	166660	166660	167206	167206	167206
Difference in Means	0.025	0.019	1.703	0.022	0.090	0.019	0.007
	(0.016)	(0.008)	(0.136)	(0.002)	(0.003)	(0.001)	(0.004)
Remote							
Before Opening							
Mean	0.099	0.035	0.742	0.032	0.032	0.045	0.278
	(0.012)	(0.008)	(0.035)	(0.002)	(0.001)	(0.0004)	(0.008)
N	76796	76796	78581	78581	82165	82165	82165
After Opening							
Mean	0.103	0.033	1.944	0.047	0.088	0.058	0.287
	(0.012)	(0.009)	(0.115)	(0.003)	(0.003)	(0.001)	(0.007)
N	50127	50127	50048	50048	50183	50183	50183
Difference in Means	0.004	-0.002	1.202	0.015	0.056	0.013	0.009
	(0.017)	(0.012)	(0.120)	(0.002)	(0.003)	(0.001)	(0.006)

Notes: Based on observations from April 15, 2020 to December 2, 2020. Standard errors that are two-way clustered on county and date are reported in parentheses.

- *"Finally I liked the graphs at the beginning, but the reader has to have some discussion either before or just after that which explains in more detail precisely what the data is and where it comes from. A table summarizing data characteristics would also be helfpful."*

Following your suggestion, we added the data section in which we explains precisely the definition of variables and their data sources. We explain the data in details before we present Figure 1 in the newly added data section. We also present a summary table in Table R1 with the following discussion:

The school opening dates spread from the beginning of August to late September across counties, where hybrid teaching is more common than remote or in-person teaching (SI Appendix, Fig. S2(k)). Reflecting the steady increase in cases from late September to November of 2020 in the U.S. (SI Appendix, Fig. S2(b)), the growth rates of cases and deaths, as well as the number of weekly confirmed cases and deaths, are higher in the period after the school opening compared to the period before. This rise in cases and deaths after the school opening is more pronounced in the counties with in-person or hybrid teaching than those with remote teaching. The K-12 school and workplace visits are also higher after the school openings than before, especially for counties with in-person and hybrid school openings. On the other hand, the mean per-device restaurant visits do not change much before and after the school opening, regardless of teaching methods.

In addition, we present the following summary tables and figure in SI Appendix:

- Table S1 presents the summary statistics for all variables used in our analysis.
- Table S2 presents correlation coefficients across outcome variables and controls.
- Table S3 reports correlation coefficients across different mitigation measures (mask mandates for staffs, mask mandates for students, sports activities, online instructions).
- Fig. S3 shows the evolutions of the 5, 25, 50, 75, and 95 percentile values of variables over time, a fraction of counties of which schools are open in-person/hybrid/remote over time, and a fraction of counties that implement three NPIs (mask mandates, ban gathering, and stay-at-home orders).

We discuss the findings from Tables S1-S3 and Fig. S3 in the data section.

Reply to the Comments on "The Association of Opening K-12 Schools and Colleges with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis"

Reply to Referee 1

Thank you very much for your helpful and constructive comments on our paper. We appreciate the time you spent in reviewing this paper. We have followed your recommendations, and those of the other reviewers', as far as possible. Major changes are:

1. Following the suggestion by the editor and the referees, we added the event-study regression analysis. Fig. 2 and 3 indicate that the gap in the average level of cases/deaths between in-person/hybrid school opening and remote school opening grows larger as more weeks pass after the school openings. We also provide the robustness checks in the panel regression analysis by estimating a specification with the log of weekly cases in place of the log difference as outcome variables.

2. Following the advice of the editor and the referee, we rearrange the order of our presentation by first discussing the data and the definition of variables in details and then providing a summary table with descriptive analysis before conducting the event-study analysis and the panel regression analysis.

3. We explicitly write a regression model which is motivated by the SIRD model, and we clearly state the unconfoundedness assumption under which we may interpret our panel regression result as causal.

4. Given the comment of the referee that "the evidence regarding colleges is substantially less strong than the evidence for K-12 schools" as well as the space constraint after adding the event-study analysis, we decided to drop our focus on the effect of college openings and mainly focus on the effect of K-12 school openings in our revised manuscript. In particular, we made the following changes.

    - We moved Figure 2 in the previous manuscript and its related discussion on the effect of college openings on cases and deaths to SI Appendix.
    - We moved Table 2(b) in the previous manuscript on the effect of opening colleges and K-12 schools on the visits to restaurants and bars to SI Appendix because Table 2(b) was presented to highlight the effect of college openings on an increase in the visits to bars.
    - The paper's title has been changed to "The Association of Opening K-12 Schools with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis" by dropping "and Colleges" from the previous title.

5. For the death growth regression, we now use the outcome variable defined by the log difference over 21 days in weekly deaths rather than the log difference over 7 days because the available evidence from the CDC shows that the time lag between infection and death reporting is stochastic and spreads over at least 2 weeks (c.f., Table 2 of `https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html`). We also use the explanatory variables that are lagged by 35 days to capture the time lag of infection and death reporting, where we provide robustness checks with respect to the choice of lags in sensitivity analysis.

6. We now use the standard fixed effects estimator without bias correction rather than debiased fixed effects estimator for the effect of school openings on full-time workplace visits, staying home devices, visits to restaurants and bars in Tables 3 and Table S4 in the revised manuscript (which corresponds to Table 2(a)(b) in the previous manuscript). This is because the specification for mobility outcome does not have lagged dependent variables in the covariates and there is no need for implementing bias-corrections. The results are similar after changing from the debiased fixed effects estimator to the standard fixed effects estimator.

Reply to Referee 1's Comments

- *Methods: "In general I found the methods used in the paper unnecessarily complicated, and also described in an opaque way. Most applied researchers would approach this problem with simple difference-in-difference style methods comparing changes in outcomes for counties where schools reopened to corresponding changes for counties that didn't reopen over the same time period (for example, this is the approach used in other papers on the spread of Covid-19 that the authors reference). The authors instead employ a more complicated dynamic panel approach without explaining the motivation for using this approach over simpler methods. In addition, there is little attention paid to explaining or validating the assumptions underlying the research design."*

  Following your suggestion, we now present the result of event-study design using the number of new confirmed cases/deaths per 1000 as outcome variables to examine how these outcome variables changes before and after the reopening of schools across different modes. Consistent with our finding from Figure 1 and the panel regression analysis, the result shows that the gap in the average number of new confirmed cases/deaths between in-person/hybrid school opening and remote school opening grows larger as more weeks pass after the school openings. We would like to thank you for encouraging us to present these results because we believe that these results reinforce our findings.

  Because different counties open their schools at different points in time, we needed to carefully consider how to implement the event-study design. As is widely discussed recently (c.f., Goodman-Bacon, 2018; Sun and Abraham, 2020), the Two-Way-Fixed-Effects (TWFE) regression with leads and lags give biased estimates under the staggered treatment design when the treatment effects are dynamic and evolve over time. Because the SIRD model suggests that the school openings and other policies may affect the growth rate of cases and deaths, the effect of school openings on the level of cases and deaths is expected to evolve overtime. Therefore, naively implementing TWFE event-study regression may result in biased estimator and, in particular, its pre-treatment estimates may indicate the presence of pre-trend even though there is no pre-trend in the true Data Generating Process. The bias arises because the pre-treatment estimate under TWFE regression partly relies on the wrong control units that have already exposed to treatment under the staggered treatment design. See the simulation results in `https://hhsievertsen.shinyapps.io/kylebutts_did_eventstudy/`, which also shows that the estimator of Callaway and Sant'Anna (2020) works very well.

Given this, we implemented the event-study design in the following two ways.

First, in the spirit of the analysis by Sun and Abraham (2020) and Callaway and Sant'Anna (2020), we divide the sample into different subsamples, where each subsample contains the observation with the similar school opening dates and then, for each of subsamples, we run the following event-study regression with weekly dummies of leads and lags for three school opening modes (i.e., in-person, hybrid, and remote) and county fixed effects but without time-fixed effects:

$$Y_{it} = \sum_{p \in \{\text{in-person,hybrid,remote}\}} \sum_{w=-8}^{22} \gamma_w^p D_{\tau,it}^p + \alpha_i + \epsilon_{it}, \qquad (2)$$

where $D_{\tau,it}^p$ takes the value equal to 1 if school has been opened for τ weeks (or will be opened after $-\tau$ weeks if $\tau < 0$) with teaching mode p in county i at day t.

Figure R3 presents the estimated coefficients for weekly dummies of leads and lags with 95 percent confidence intervals for the number of new weekly confirmed cases per 1000, that of deaths per 1000, and K-12 school visits per device as outcome variables for three subsamples: the first subsample uses the county observations which opened schools before Aug 23, 2020, the second consists of the counties that opened schools between Aug 24 and Sept 6, and the third uses the counties with the school opening dates after Sept 7, 2020. Dividing even finer subsamples within a week of school opening dates give similar results although the results look much noisier given the smaller sample size. Across all three subsamples for both cases and deaths, there is no clear violation of parallel trend in pre-treatment period for the last three weeks before school openings. The results also clearly illustrate that the gap in weekly cases/deaths per 1000 between remote opening and in-person/hybrid opening grow over time after the school opening date for all three subsamples.

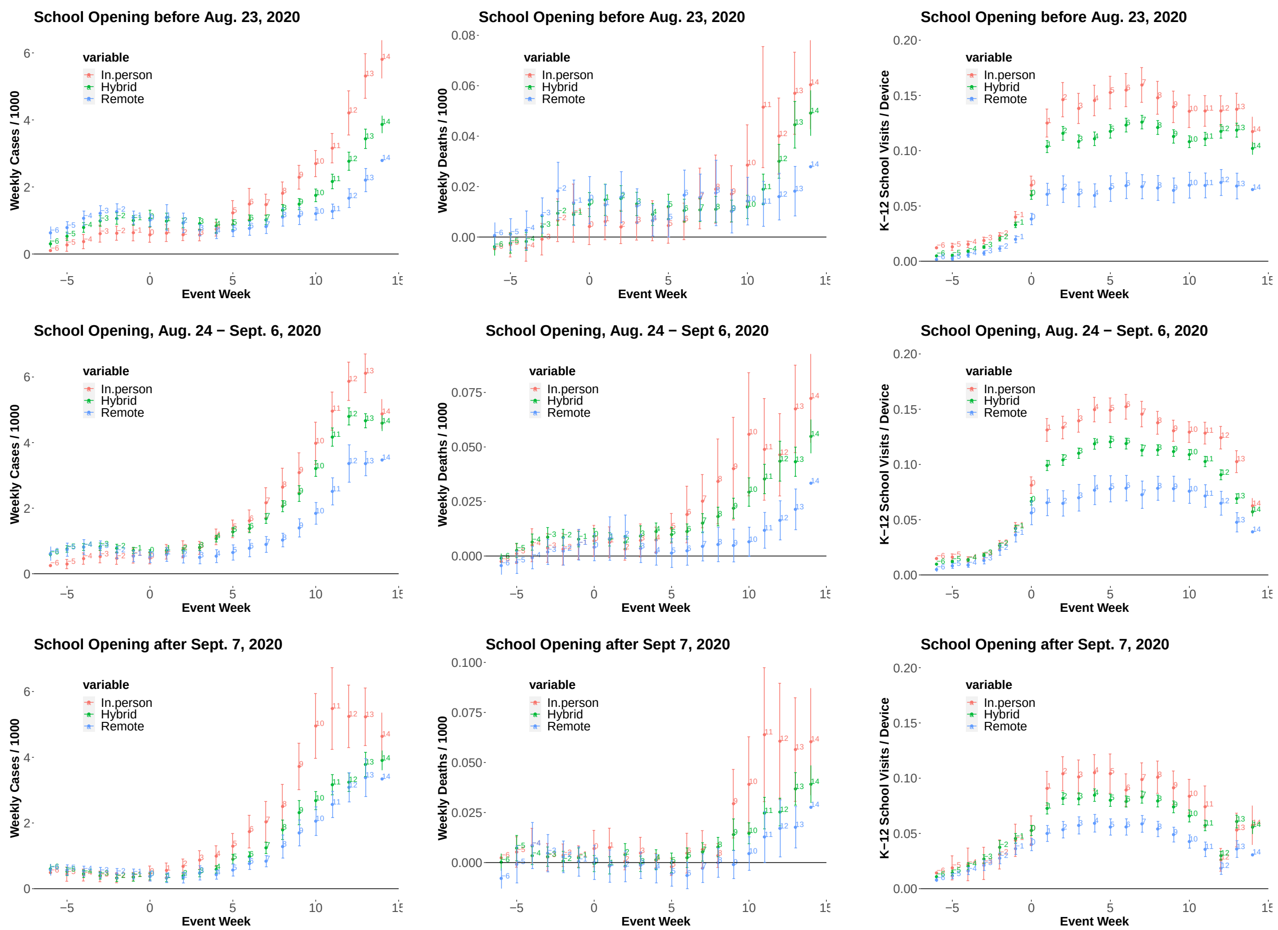
Second, we follow the proposed method of Callaway and Sant'Anna (2020) using their did R package to estimate the difference in the dynamic treatment effects between remote school openings and in-person/hybrid school openings. To estimate the group-time specific average treatment effect, where the group is defined by a set of counties with the same school opening date, their method only uses the "never-treated-units" and "not-yet-treated-units" as the controls while excluding the "already-treated-units" from the control group, where we take the counties with remote opening plan as the "never-treated-units." The parameter of interest here is the average of the group-time specific average treatment effects of in-person or hybrid opening against remote opening across groups with different school opening dates.

As shown in Figure R4, the result of Callaway and Sant'Anna (2020) estimator is consistent with the findings based on the event-study regression in Figure R3. Figure R4(a)-(d) shows that the gap in school visits between counties with remote opening plan and those with in-person/hybrid opening plans increases after the school opening date and these increases are especially pronounced for counties without any mask mandate for staffs. All estimates in pre-treatment periods are not significantly different from zero. Figure R4(e)-(p) report the estimated difference in weekly deaths and the log of weekly cases and deaths between counties with remote opening and those with in-person/hybrid opening gradually increases after the school opening date, especially for counties with no mask mandates at school. Figure R4(q)-(x) statistically confirm the patterns reported in Figure 1(c)(d) in the main text, showing that K-12 school visits and visits to full-time workplaces increase after the opening of schools for counties with in-person/hybrid teaching relative to those with remote teaching.

Finally, Figure R5 show that there is no clear evidence for the association of the school opening date across different teaching modes and mask mandates with the number of visits to restaurants, bars, gyms, and churches, especially in the pre-treatment periods. This suggests that, on average, other possibly unobserved county-level confounders (e.g., lockdown policies and changes in people's behavior) may not be systematically related to the timing of school opening, teaching modes, and mitigation measures at school. There are a couple of exceptions. First, the positive estimate with a large confidence interval at event week 0 in Figure R5(e)(f) may suggest a possibility that teachers and parents went out for bars on the school opening week in some counties that opened school in-person—one of the possible mechanisms through which in-person school openings affected the subsequent case growth although the effect at event week 0 is imprecisely estimated. Second, college visits appear to increase after K-12 school openings, which may reflect that the timing of school openings coincide between K-12 schools and colleges/universities within a county. In our panel regression analysis, we always control for college visits while the robustness of the result is examined with respect to adding the control of restaurants, bars, gyms, and churches in our sensitivity analysis.

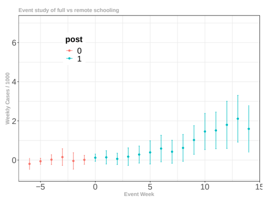
Figure R3: The event-study regression estimates before and after the opening of K-12 schools

(a) Cases by Opening Modes **(b) Deaths by Opening Modes** **(c) K-12 Visits by Opening Modes**

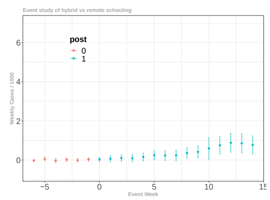


Notes: The figures plot the estimated coefficients for weekly dummies of leads and lags in the regression specification (2) with 95 percent confidence intervalsfor cases per 1000, deaths per 1000, and K-12 school visits per device as outcome variables for three subsamples classified by school opening dates.

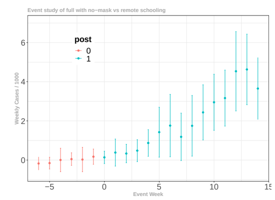
Figure R4: The average treatment estimates obtained using the DID method from Callway and Sant'Anna (2020)



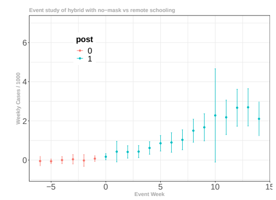
(a) Cases: In-person vs. Remote



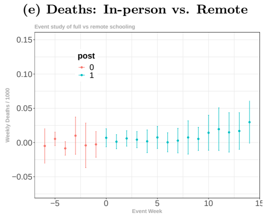
(b) Cases: Hybrid vs. Remote



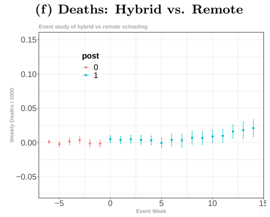
(c) Cases: In-person/No-Mask vs. Remote



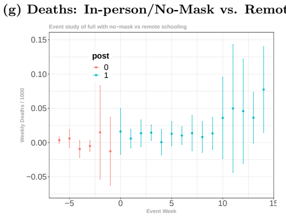
(d) Cases: Hybrid/No-Mask vs. Remote



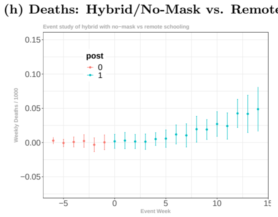
(e) Deaths: In-person vs. Remote



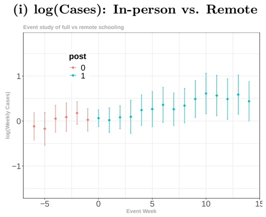
(f) Deaths: Hybrid vs. Remote



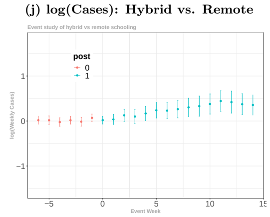
(g) Deaths: In-person/No-Mask vs. Remote



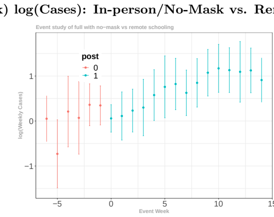
(h) Deaths: Hybrid/No-Mask vs. Remote



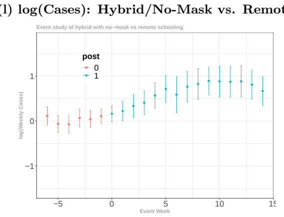
(i) log(Cases): In-person vs. Remote



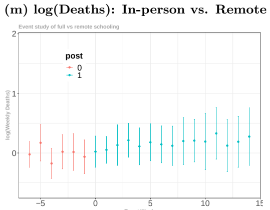
(j) log(Cases): Hybrid vs. Remote



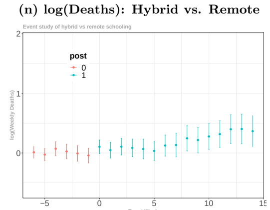
(k) log(Cases): In-person/No-Mask vs. Remote



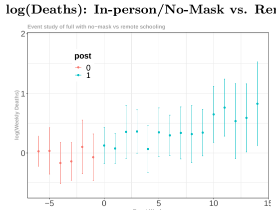
(l) log(Cases): Hybrid/No-Mask vs. Remote



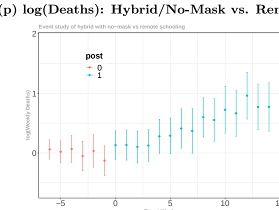
(m) log(Deaths): In-person vs. Remote



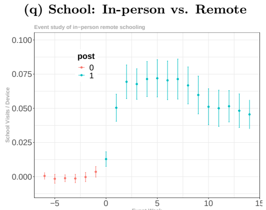
(n) log(Deaths): Hybrid vs. Remote



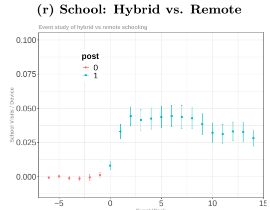
(o) log(Deaths): In-person/No-Mask vs. Remote



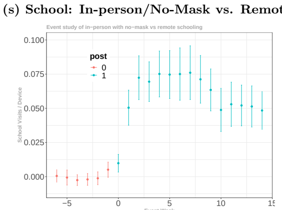
(p) log(Deaths): Hybrid/No-Mask vs. Remote



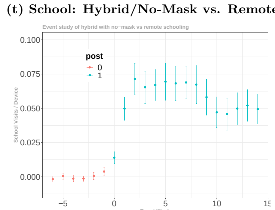
(q) School: In-person vs. Remote



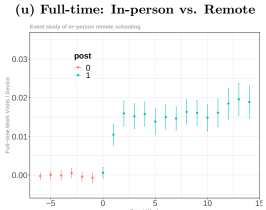
(r) School: Hybrid vs. Remote



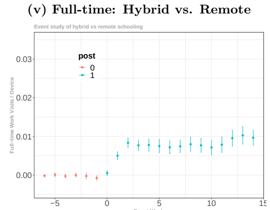
(s) School: In-person/No-Mask vs. Remote



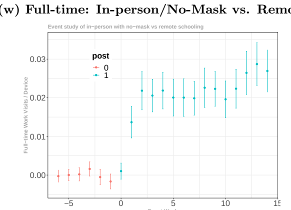
(t) School: Hybrid/No-Mask vs. Remote



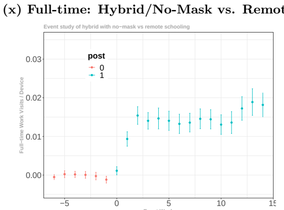
(u) Full-time: In-person vs. Remote



(v) Full-time: Hybrid vs. Remote



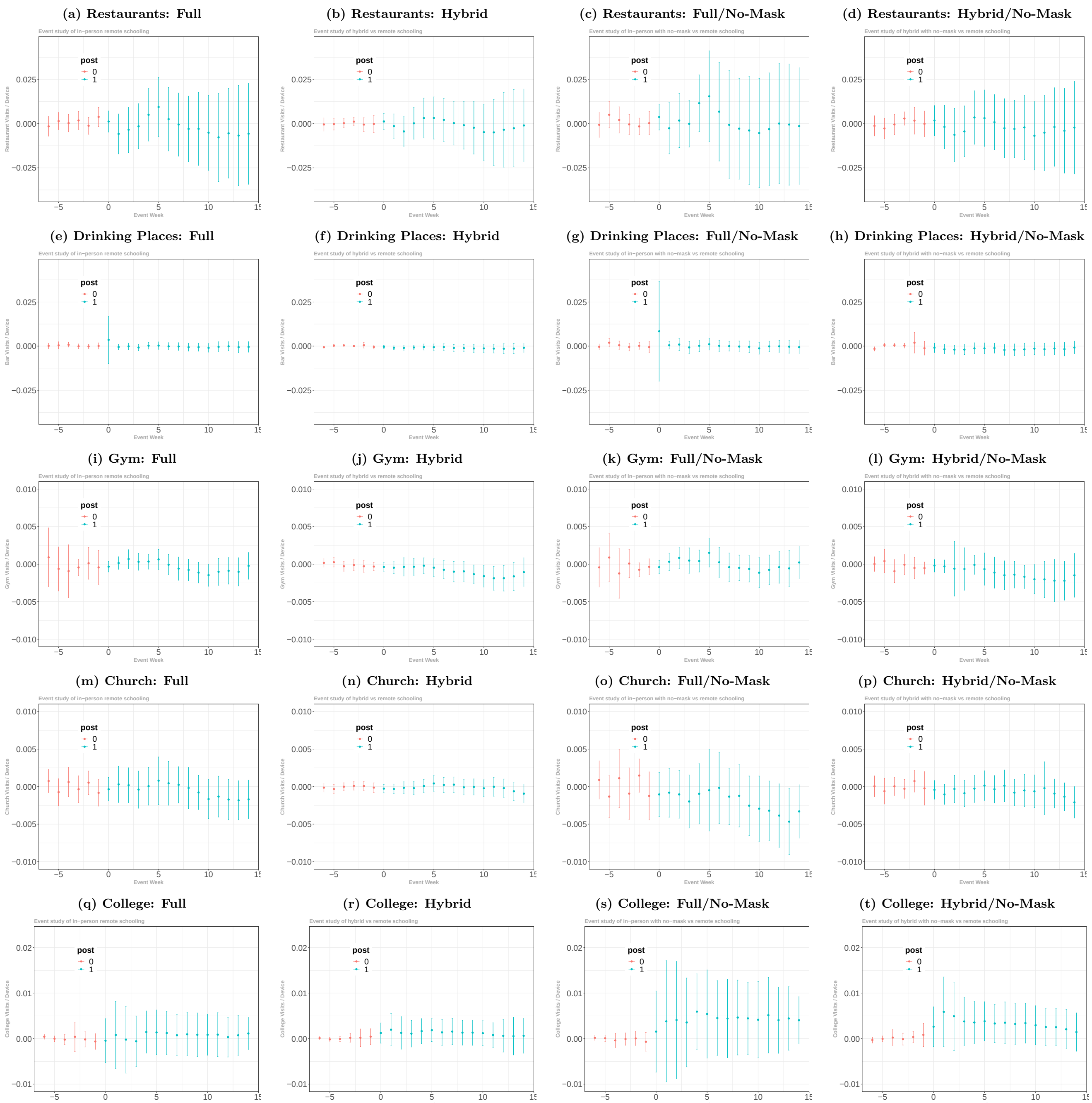
(w) Full-time: In-person/No-Mask vs. Remote



(x) Full-time: Hybrid/No-Mask vs. Remote

Notes: (a) plots the estimates of the average dynamic treatment effect of in-person openings relative to the counties with remote openings as well as the counties that have not opened yet on cases per 1000 using a subset of counties with either in-person opening or remote-opening, where we use the estimation method of Callaway and Sant'Anna (2020) implemented by their did R package. Similarly, (b), (c), and (d) plots the estimates of the average dynamic treatment effect of school opening with hybrid, in-persion/mask mandates, and hybrid/mask mandates teaching methods, respectively, using a subset of counties with the corresponding teaching method as well as remote-opening. (e)-(h), (i)-(l), (m)-(p), (q)-(t), and (u)-(x) report the estimates of the average dynamic treatment effect on deaths per 1000, log(cases), log(deths), per-devise visits to K-12 schools, and per-devise visits to full-time workplaces, respectively.

Figure R5: The event-study regression estimates for behavior variables obtained using the DID method from Callway and Sant'Anna (2020)



(a) Restaurants: Full (b) Restaurants: Hybrid (c) Restaurants: Full/No-Mask (d) Restaurants: Hybrid/No-Mask

(e) Drinking Places: Full (f) Drinking Places: Hybrid (g) Drinking Places: Full/No-Mask (h) Drinking Places: Hybrid/No-Mask

(i) Gym: Full (j) Gym: Hybrid (k) Gym: Full/No-Mask (l) Gym: Hybrid/No-Mask

(m) Church: Full (n) Church: Hybrid (o) Church: Full/No-Mask (p) Church: Hybrid/No-Mask

(q) College: Full (r) College: Hybrid (s) College: Full/No-Mask (t) College: Hybrid/No-Mask

- *"The authors fit dynamic panel models that relate the log change in cases over two week periods to county fixed effects, state-by-time effects, and lagged levels of cases. There are a number of issues to consider here: Why estimate the model in first differences rather than* *levels? A fixed effects model in first differences is implicitly a model with county trends. I would have liked to see the simpler model with county intercepts before adding trends."*

Given your comment, we now motivate our use of the weekly growth rate of cases—approximated by the first difference in the log of weekly cases—as an outcome variable in our regression specification by deriving our regression specification from the Susceptible-Infectious-Recovered-Deceased (SIRD) model in the main text. In particular, our specification follows from the SIRD model by relating the growth rate of cases to the rate of infection spread approximated by a linear function of observable policies, the lagged log of cases, and other confounders.

On the other hand, because the lagged log of cases are included as covariates, our regression model can be interpreted as a model with county intercepts rather than a model with county trends. To see this, consider a slightly modified version of the model with the log difference in cases over two weeks as outcome variable:

$$\begin{aligned}\log Case_{it} - \log Case_{i,t-14} = \beta' Visit_{i,t-14} + \sum_{\tau=14,21,28} \beta_{y,\tau} \log Case_{i,t-\tau} \\ + \gamma' NPI_{i,t-14} + \theta\, Test_{it} + \alpha_i + \delta_{s(i),w(t)} + \epsilon_{it}, \qquad (3)\end{aligned}$$

The regression model (3) can be re-written as:

$$\begin{aligned}\log Case_{it} = \beta' Visit_{i,t-14} + \sum_{\tau=14,21,28} \tilde{\beta}_{y,\tau} \log Case_{i,t-\tau} \\ + \gamma' NPI_{i,t-14} + \theta\, Test_{it} + \alpha_i + \delta_{s(i),w(t)} + \epsilon_{it}, \qquad (4)\end{aligned}$$

where $\tilde{\beta}_{y,14} = 1 + \beta_{y,14}$ and $\tilde{\beta}_{y,\tau} = \beta_{y,\tau}$ for $\tau = 21, 28$.

Given your suggestion of estimating a model with county intercepts, we estimated specification (4) together with (3) in Tables R2-R3. Both results are qualitatively similar to those in Table 1 in the main text, suggesting that an increase in visits to schools and opening schools in-person are positively associated with an increase in cases after two weeks conditional on the lagged cases. The interaction terms of school visits or in-person school openings with no-mask mandates for staffs are positively estimated, possibly indicating the role of mitigation measures for reducing the impact of school opening in-person on the transmission of COVID-19 at schools. The estimated magnitudes in Tables R2-R3 are generally larger than the estimates in Table 1 in the main text because the outcome variable in specification (3) approximate the growth rate of cases over two weeks rather that the weekly growth rate.

We included Table R3 in SI Appendix and briefly discuss the robustness of our panel regression results when we use the log of weekly cases rather than the log difference as outcome variable.

Table R2: The Association of School/College Openings and NPIs with log of Cases in the United States: Debiased Estimator

	Dep. Variable: log(Cases)$_t$- log(Cases)$_{t-14}$			
	(1)	(2)	(3)	(4)
College Visits, 14d lag	0.226	−0.346	0.205	−0.328
	(0.240)	(0.260)	(0.230)	(0.252)
K-12 Visits, 14d lag	1.185***	0.469***		
	(0.142)	(0.155)		
K-12 Visits × No-Mask, 14d lag		1.324***		
		(0.176)		
K-12 In-person, 14d lag			0.241***	0.115***
			(0.041)	(0.041)
K-12 Hybrid, 14d lag			−0.023	−0.139***
			(0.028)	(0.029)
K-12 Remote, 14d lag			−0.337***	−0.369***
			(0.028)	(0.032)
K-12 In-person × No-Mask, 14d lag				0.115*
				(0.059)
K-12 Hybrid × No-Mask, 14d lag				0.174***
				(0.040)
Mandatory mask, 14d lag	−0.519***	−0.536***	−0.507***	−0.512***
	(0.043)	(0.041)	(0.043)	(0.041)
Ban gatherings, 14d lag	−0.403***	−0.405***	−0.427***	−0.415***
	(0.046)	(0.069)	(0.055)	(0.070)
Stay at home, 14d lag	−0.294***	−0.273**	−0.269***	−0.262**
	(0.095)	(0.114)	(0.104)	(0.123)
log(Weekly Cases), 14d lag	−0.231***	−0.220***	−0.220***	−0.212***
	(0.012)	(0.010)	(0.012)	(0.010)
log(Weekly Cases), 21d lag	−0.457***	−0.456***	−0.453***	−0.453***
	(0.007)	(0.007)	(0.007)	(0.007)
log(Weekly Cases), 28d lag	−0.009	−0.005	−0.004	−0.003
	(0.006)	(0.006)	(0.006)	(0.006)
Test Growth Rates	0.001	0.0005	0.001	0.001
	(0.001)	(0.001)	(0.001)	(0.001)
County Dummies	Yes	Yes	Yes	Yes
State× Week Dummies	Yes	Yes	Yes	Yes
Observations	668,101	527,576	593,265	511,939
R^2	0.375	0.376	0.372	0.377

Notes: Dependent variable is the difference in the log of past 7 days cases between day t and day $t-14$. Regressors are 7-days moving averages of corresponding daily variables and lagged by 2 weeks to reflect the time between infection and case reporting except that we don't take any lag for the log difference in test growth rates. Because the 2 weeks lagged log cases variable is in the control, the estimated coefficients can be interpreted as the covariate's association with case growth over 2 weeks. All regression specifications include county fixed effects and state-week fixed effects to control for any unobserved county-level factors and time-varying state-level factors such as various state-level policies. The debiased fixed effects estimator is applied. The results from the estimator without bias correction is presented in SI Appendix, Table S1. Asymptotic clustered standard errors at the state level are reported in bracket. *p<0.1; **p<0.05; ***p<0.01

- *"The authors include lags of the dependent variable on the right, making this a dynamic*

Table R3: The Association of School/College Openings and NPIs with log of Cases in the United States: Debiased Estimator

	Dependent variable: log(Weekly Cases)			
	(1)	(2)	(3)	(4)
College Visits, 14d lag	1.927***	2.016***	1.945***	1.879***
	(0.253)	(0.264)	(0.233)	(0.254)
K-12 Visits, 14d lag	1.402***	0.978***		
	(0.193)	(0.222)		
K-12 Visits × No-Mask, 14d lag		0.792***		
		(0.194)		
K-12 In-person, 14d lag			0.338***	0.303***
			(0.046)	(0.048)
K-12 Hybrid, 14d lag			0.150***	0.104***
			(0.029)	(0.031)
K-12 Remote, 14d lag			−0.013	0.009
			(0.038)	(0.044)
K-12 In-person × No-Mask, 14d lag				0.059
				(0.057)
K-12 Hybrid × No-Mask, 14d lag				0.167***
				(0.048)
Mandatory mask, 14d lag	−0.278***	−0.244***	−0.270***	−0.256***
	(0.048)	(0.051)	(0.050)	(0.052)
Ban gatherings, 14d lag	−0.112	−0.067	−0.113	−0.055
	(0.155)	(0.151)	(0.155)	(0.149)
Stay at home, 14d lag	0.412***	0.435***	0.464***	0.469***
	(0.066)	(0.073)	(0.071)	(0.079)
log(Weekly Cases), 14d lag	0.408***	0.405***	0.409***	0.402***
	(0.010)	(0.009)	(0.010)	(0.009)
log(Weekly Cases), 21d lag	0.133***	0.135***	0.133***	0.134***
	(0.005)	(0.004)	(0.005)	(0.005)
log(Weekly Cases), 28d lag	0.025***	0.027***	0.023***	0.025***
	(0.006)	(0.005)	(0.006)	(0.006)
Test Growth Rates	0.005**	0.004*	0.005**	0.004*
	(0.002)	(0.002)	(0.002)	(0.002)
County Dummies	Yes	Yes	Yes	Yes
State× Week Dummies	Yes	Yes	Yes	Yes
Observations	760,422	600,958	675,405	583,119
R^2	0.867	0.862	0.866	0.861

Notes: Dependent variable is the log of weekly positive cases. Regressors are 7-days moving averages of corresponding daily variables and lagged by 2 weeks to reflect the time between infection and case reporting except that we don't take any lag for the log difference in test growth rates. Because the 2 weeks lagged log cases variable is in the control, the estimated coefficients can be interpreted as the covariate's association with case growth over 2 weeks. All regression specifications include county fixed effects and state-week fixed effects to control for any unobserved county-level factors and time-varying state-level factors such as various state-level policies. The debiased fixed effects estimator is applied. The results from the estimator without bias correction is presented in SI Appendix, Table S1. Asymptotic clustered standard errors at the state level are reported in bracket. *p<0.1; **p<0.05; ***p<0.01

panel specification. Again, why not start with simpler models without the lagged dependent variable controls? Is there evidence that these controls are necessary to eliminate differential pre-trends across areas with different reopening policies?"

Our dynamic panel data analysis is based on a different research design from the diff-in-diff research design, where our identification assumption is the conditional random assignment of policies given the controls. The diff-in-diff replies on parallel trend assumption but its validity is sensitive to the transformation of outcome variable, where the only way to rationalize this assumption such that it holds for any transformation of outcome variable is to assume random assignment (Roth and Sant'Anna, 2021). Therefore, the conditional random assignment, or so-called the unconfoundedness assumption, is the crucial assumption.

With this clarification, we believe that controlling for the lagged cases or deaths is critical to satisfy the uncounfoundedness assumption for our panel regression research design (i.e., the conditional randomness of the timing/mode of school openings given observed confounders) because the timing and the mode of school openings are very likely to be affected by the number of lagged cases or deaths. In fact, the decision to reopen schools in California and Oregon depended on trends in local case counts as discussed in Goldhaber-Fiebert, Studdert, and Mello (2020).

Furthermore, the estimated coefficients of the lagged log cases or deaths are always significant and substantial with predicted signs, which are consistent with the hypothesis that people change their risk-taking behavior in response to new information on the confirmed cases. Because the lagged log cases are correlated with school opening timings/modes, and other policy variables, not controlling for the lagged log cases will result in the omitted variable bias.

Finally, our regression specification is derived from the well-known epidemiological model, i.e., the SIRD model, which helps us to think about how the spread of COVID-19 happens. The lagged dependent variables are obvious confounders once we realize that people voluntarily change their behavior (mask wearings, social distancing, washing hands) in response to the risk of virus transmission. People behave more cautiously when the reported case/death number is high, which will reduce the subsequent case or death growth even without policy interventions. As shown in Table 3 of the revised manuscript, the current and the past log cases are strongly correlated with the proportion of staying home devices even after controlling for county-level lockdown policies as well as state-week fixed effects, suggesting that people are more likely to stay home when the transmission risk is high.

For these reasons, we keep a dynamic panel specification with the lagged dependent variable as our baseline specification.

In response to your comment, we estimated the following regression specification which is a model without the lagged dependent variable controls:

$$\log Case_{it} = \beta' Visit_{i,t-14} + \gamma' NPI_{i,t-14} + \theta Test_{it} + \alpha_i + \delta_{s(i),w(t)} + \epsilon_{it}. \tag{5}$$

Given that there is no lagged dependent variable control, we estimate (5) using the standard fixed effects estimator without debiasing. Table R4 reports the result, indicating that in-person/hybrid school openings is associated with the higher number of new confirmed cases especially when there is no mask mandate for staff.

We included Table R4 in SI Appendix and briefly discuss the robustness of our panel regression results in the main text at the end of "Results" section.

- *"The authors employ a "split-panel" approach that corrects for bias in fixed effect estimates of lagged dependent variable specifications in short panels. There are many other approaches to this problem, for example methods based on instrumenting with longer lags (e.g. Anderson-Hsiao, Arellano-Bond, or system GMM). Estimates in dynamic panel models can be extremely sensitive to which method is chosen so it would be good to see comparisons to other approaches."*

  The key additional moment condition for system GMM in Blundell and Bond (1998) is derived under the assumption that the initial observations are drawn from the stationary distribution, which is justified if the stationary process has been run for a long time before we observe the first observation in our data set. We don't think that this is a reasonable assumption for the stochastic process for cases and deaths for COVID-19. Therefore, we don't think the use of system GMM is justifiable.

  An important issue for using the method of Anderson-Hsiao and Arellano-Bond is that, because the effective length of panel dimension is large (i.e., more than 24×7=168 days), there are a very large number of moments we need to consider, which also necessitates bias correction (e.g., by sample splitting) as discussed in Chen, Chernozhukov, and Fernández-Val (2019).

  Specifically, when the number of time periods is $T = 27 \times 7 = 168$, Arellano-Bond estimator uses a very large number of moment conditions with the order of $m = O(T^2) = 28224$ to identify the parameter. The cross-sectional number of observations is $n = 3144$, at most. For the asymptotic statistical inference without bias correction to be valid, it is required that $m^2/(n \times T) \to 0$ but in our case, $m^2/(n \times T) = 1508$, which is far away from 0.

  In short, Arellano-Bond estimator is biased and requires bias correction in our context (Chen, Chernozhukov, and Fernández-Val, 2019). Furthermore, even after bias correction, Arellano-Bond estimator is considerably less efficient than the bias corrected fixed effects estimator.

Table R4: The Association of School/College Openings and NPIs with log of Cases in the United States: Standard Fixed Effects Estimator

	Dependent variable: log(Weekly Cases)			
	(1)	(2)	(3)	(4)
College Visits, 14d lag	1.871***	1.964***	1.972***	1.916***
	(0.398)	(0.407)	(0.376)	(0.398)
K-12 Visits, 14d lag	0.955***	0.640**		
	(0.232)	(0.297)		
K-12 Visits × No-Mask, 14d lag		0.616**		
		(0.290)		
K-12 In-person, 14d lag			0.368***	0.332***
			(0.068)	(0.079)
K-12 Hybrid, 14d lag			0.184***	0.138***
			(0.047)	(0.046)
K-12 Remote, 14d lag			−0.007	0.009
			(0.053)	(0.059)
K-12 In-person × No-Mask, 14d lag				0.052
				(0.102)
K-12 Hybrid × No-Mask, 14d lag				0.186***
				(0.069)
Stay at home, 14d lag	−0.093*	−0.068	−0.097*	−0.072
	(0.055)	(0.055)	(0.056)	(0.055)
Ban gatherings, 14d lag	−0.167*	−0.181*	−0.193**	−0.186*
	(0.094)	(0.106)	(0.093)	(0.103)
Stay at home, 14d lag	0.038	0.054	0.060	0.081
	(0.085)	(0.097)	(0.092)	(0.103)
Test Growth Rates	0.006**	0.005**	0.006***	0.005**
	(0.002)	(0.002)	(0.002)	(0.002)
County Dummies	Yes	Yes	Yes	Yes
State× Week Dummies	Yes	Yes	Yes	Yes
Observations	760,422	600,958	675,405	583,119
R^2	0.867	0.862	0.866	0.861

Notes: Dependent variable is the log of weekly positive cases. Regressors are 7-days moving averages of corresponding daily variables and lagged by 2 weeks to reflect the time between infection and case reporting except that we don't take any lag for the log difference in test growth rates. All regression specifications include county fixed effects and state-week fixed effects to control for any unobserved county-level factors and time-varying state-level factors such as various state-level policies. Asymptotic clustered standard errors at the state level are reported in bracket. *p<0.1; **p<0.05; ***p<0.01

Given these reasons, we decided not to pursue system GMM, Anderson-Hsiao, and Arellano-Bond approaches.

We present the fixed effects estimates without bias correction, showing qualitatively similar results to the debiased fixed effects estimates.

- *"In a basic diff-in-diff or event study design it is standard to spend considerable effort presenting transparent tests to convince the reader that the underlying parallel trends assumption is satisfied. The authors spend little attention paid to whether the school openings underlying the authors' estimates look like a "clean" experiment, and whether the controls they include help to eliminate potential sources of bias."*

  Our research design for dynamic panel regression analysis is not diff-in-diff but we do now provide event-study analysis. In our event-study analysis, we find no clear violation of the parallel trend in pre-treatment periods.

  On the other hand, our main research design for dynamic panel regression analysis is fundamentally different from the diff-in-diff or the event study design. The key assumption to interpret our estimate as causal for our panel data analysis is the uncounfoundedness assumption that the timing and the teaching mode of school opening is random conditional on our specified confounders (county fixed effects, state-week fixed effects, observed policy variables such as mask mandates, stay-at-home, and ban gathering). While the assumption cannot be tested, we provide a number of sensitivity analysis (e.g., additionally conditioning on observed mobility variables such as visits to restaurants, bars, gyms, churches with various lag length and changing the lag length), suggesting that the results are robust. Furthermore, we explicitly state the assumption and the limitation as follows:

  > First, our study is observational and, therefore, should be interpreted with great caution. It only has a causal interpretation in a structural model under the unconfoundedness assumptions that might not hold in reality. While we present sensitivity analysis with various controls, including county dummies and interactions of state dummies and week dummies, the decisions to open K-12 schools may be endogenous and correlated with other unobserved time-varying county-level factors that affect the spread of COVID-19. For example, people's attitudes toward social distancing, hand-washing, and mask-wearing may change over time (which we cannot observe in the data). Their changes may be correlated with school opening decisions beyond the controls we added to our regression specifications.

- *"There is not much discussion of the data and no table of descriptive statistics. It would be good to give the reader a better sense of the structure and content of the data."*

Given the constraint on page limits as well as the limits to the number of tables and figures, the previous manuscript presented descriptive statistics in SI Appendix. We now present the summary table that highlights how the means of case/death growth, case/death per 1000, visits to K-12 schools, workplaces, and restaurants change before and after school openings across three different teaching methods in the main text and explain the data in more detail before we present Figure 1 in the paper.

- *Exposition: "The authors do not explain the data sources or methods they use until the end of the paper after all the results have been presented. This makes it hard to understand the earlier parts of the paper. It would be good to lay out the data sources, research design, and methods before discussing the findings."*

  In our previous version, we followed the suggested PNAS format which puts "Materials and Methods" at the very last section of the paper. Given your request, we now discuss the data and the methods before discussing the findings.

- *"I did not find the causal graph diagram useful for understanding the research design or underlying assumptions. It would be clearer to spell out what the authors need to assume to give the results a causal interpretation, and investigate the relevant assumptions."*

  The causal graph is the standard way to communicate unconfoundedness research design in epidemiology (see Hernán and Robins, 2020).

  When we presented our paper to different audiences and communicated with other researchers including those in epidimiology, we often received positive feedback on our use of the causal graph diagram and some audience think that the causal graph diagram is helpful to understand the relevant assumptions and potential mechanisms through which K-12 school openings may affect the spread of COVID-19. The causal graph diagram clarifies our interpretation of covariates such as the lagged log cases/deaths, i.e., information based on which people changes their behavior. Our panel regression analysis on the association between school openings and mobility is also motivated by the mechanism highlighted by the causal graph diagram, where mobility plays a role of mediator through which the school openings affect the spread of COVID-19. Therefore, we would like to keep the causal graph diagram as it is.

  Following your suggestion, we now explicitly derive the regression specification in equations [5]-[6] and clearly state the assumptions under which the results may be interpreted as causal in the main text as follows:

  > For parameter identification, we assume that the error ϵ_{it} in [5] is orthogonal to the observed explanatory variables of school visits/openings, NPIs, test rates, and the past log cases, county fixed effects, and state-week fixed effects.

> The estimated parameters for school openings can be causally interpreted under the unconfoundedness assumption that the variables related to school openings (school visits, opening dates, teaching methods) are as good as randomly assigned after conditioning on other controls, county fixed effects, and state-week fixed effects.

The details for our research design are provided in SI Appendix, The Model and Methods.

- *Policy implications:'As the authors note, several other studies have used panel research designs to study the impacts of school closures and reopenings on the spread of Covid-19. These other studies tend to use simpler diff-in-diff and event study models and find little or no impact on spread. It's hard to know what to make of the contrast between these studies and the authors' results. Is this due to a difference in methods, contexts, or both?" "State et al. (2020), Isphording et al. (2020), and Borusyak et al. (2020) study school closures and reopenings in Europe and generally find no effect on Covid cases. It is worth connecting to these studies."*

  We think both a difference in context and a difference in methods contributes to the difference in findings.

  First, in a recently paper published at *Science*, Lessler et al. (2021) find evidence for the association between in-person schooling and COVID-19-related outcomes across counties in the United States although they use a very different data set from ours because their data is based on a massive online survey. Our paper becomes publicly available and submitted to PNAS on February 23, 2021 while their paper becomes publicly available on March 1, 2021. They also find the importance of school-based mitigation measures, such as mandating mask-wearing at schools, to reduce the risks. Their results are overall consistent with our findings.

  We suspect that a variation across counties in the United States is larger than a geographical variation in European countries and, perhaps, more appropriate mitigation measures had been implemented in European countries when they opened schools than US counties that opened schools in-person without mitigation measures such as mask mandates for staffs. For example, Isphording et al. (2021) conclude that "school re-openings in Germany under strict hygiene measures combined with quarantine and containment measures have not increased the number of newly confirmed SARS-CoV-2 infections," suggesting the potential importance of mitigation measures.

  In addition, the cross-sectional sample size in our study (3000+ counties) is larger than the cross-sectional sample size of Isphording et al. (2020) and Borusyak et al. (2020) (for example, in Germany study of Isphording et al. (2020), there are 401 counties). The larger sample size in our study helps us to detect the effect of school openings on cases by increasing the power of test.

Another important source of the difference is the event window length. Isphording et al. (2021) consider an event window of two weeks before and four weeks after. In the natural experiment of school opening in Germany, all schools eventually opened and an event window of the four weeks after school openings is the maximum window length they can consider. In our view, their post-event window of four weeks is too short.

On the other hand, in our newly-added event study, because we take a set of counties that opened schools remotely as a part of the control group (in addition to counties that are "not-yet-treated") and compare them with a set of counties that fully/hybridly opened their schools, we can estimate the difference in the effect of school openings across different model of school openings longer than 4 weeks. As illustrated in Figure R3, when the number of newly confirmed weekly cases per 1000 is used as outcome variables (which is similar to the outcome variable in both Isphording et al. (2021) and Borusyak et al. (2020)), the effect of school openings becomes more visible after 4 weeks although the difference in the trend between remote opening and in-person/hybrid opening counties starts gradually arising after 2-3 weeks.

If the school opening affects the growth rate of cases only after 2 weeks, then 3-4 weeks of the event window may be too short to clearly see the effect of school openings on the level of cases. This is especially true if cases among children are difficult to be detected and confirmed because children are more likely to be asymptomatic when they are infected. The secondary infections from children infected at schools to their parents are more likely detected and confirmed, which may contribute to the delay in the rise in reported cases after school openings. For example, Vlachos (2021) recently published in PNAS provides evidence that exposure to open schools results in an increase in infections among parents in Sweden.

We apologize, but we were not able to figure out which paper "State et al. (2020)" represents even though we have done an extensive search over internet.

Following your suggestion, we added a paragraph at the end of "Event-Study Analysis" on this:

> Our finding is consistent with Lessler et al. (2021) who examine the data from a massive online survey and find the association between in-person schooling and COVID-19-related outcomes across counties and the importance of school-based mitigation measures for reducing transmission risks in the United States. In contrast, using an event-study design, Isphording, Lipfert, and Pestel (2021) and von Bismarck-Osten, Borusyak, and Schönberg (2020) find no evidence that

> fully opening schools increased case number within the 3-4 weeks of school openings in Germany. One possible reason for these contradictory findings is that the mitigation measures in German schools may have been more effective in containing in-school transmissions than the measures adopted by the US schools with in-person openings. Another important source of the difference is that the event window length of 3-4 weeks in Isphording, Lipfert, and Pestel (2021) and von Bismarck-Osten, Borusyak, and Schönberg (2020) may be too short to identify the effect of school openings on the confirmed cases because asymptomatic, undetected cases are prevalent among children (Leidman et al., 2021).

- *"There is also not much discussion of policy implications of the authors' estimates. On the one hand, the authors find that school openings may increase the spread of Covid-19. On the other hand, school closures may lead to significant harm for students. How should we balance these competing objectives? A full cost benefit analysis is outside the scope of the paper, but it would be helpful to get a rough sense of how the costs implied by the authors' estimates of increased spread compare to the costs of closures estimated from other sources."*

  Our main actionable policy implication is to vaccinate teachers and to implement strict mitigation measures when schools are open in-person. We believe that these two actions are low-cost relative to their benefits. We now emphasize this point in abstract, significance statement, introduction as well as when we discuss the cost and the benefit of school closures.

  We believe that this policy recommendation continues to be timely, since many countries (e.g., Brazil, Argentina, India, Russia) continue to have low vaccination rates and have ongoing pandemic in force. This is especially so given that the variant of concerns with high transmission rates (the Delta variant) may become a dominant source of coronavirus infections.

  Our paper *does not* advocate closing schools and we are not keen on giving the cost and the benefit analysis of school closures because such quantitative comparison is very difficult given our current knowledge on the cost and the benefit of school closures. We believe any quantitative discussion on the cost and the benefit of school closures at this moment is deemed to be imperfect, even a rough sense, and potentially harmful because we will miss many possible sources of the cost of school closures.

  To incorporate your comment, we added several relevant papers on the cost of school closures and emphasize actionable policy implications in the last section as follows:

  > Finally, our result *does not* imply that K-12 schools should be closed. Closing schools have negative impacts on children's learning (Engzell, Frey, and Verha-

gen, 2021) and may cause declining physical and mental healths among children and their parents (Takaku and Yokoyama, 2021; Ford, John, and Gunnell, 2021; Gadermann et al., 2021). On the other hand, there is emerging evidence of long-term harm on children's health induced by COVID (Parcha et al., 2021). The decision to open or close K-12 schools requires careful assessments of the cost and the benefit by policymakers. However, given their relatively low implementation costs, our findings strongly support policies that enforce masking and other precautionary actions at school and prioritizing vaccines for education workers and elderly parents/grandparents.

Reply to the Comments on "The Association of Opening K-12 Schools and Colleges with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis"

Reply to Referee 2

Thank you very much for your helpful and constructive comments on our paper. We appreciate the time you spent in reviewing this paper. We have followed your recommendations, and those of the other reviewers', as far as possible. Major changes are:

1. Following the suggestion by the editor and the referees, we added the event-study regression analysis. Fig. 2 and 3 indicate that the gap in the average level of cases/deaths between in-person/hybrid school opening and remote school opening grows larger as more weeks pass after the school openings. We also provide the robustness checks in the panel regression analysis by estimating a specification with the log of weekly cases in place of the log difference as outcome variables.

2. Following the advice of the editor and the referee, we rearrange the order of our presentation by first discussing the data and the definition of variables in details and then providing a summary table with descriptive analysis before conducting the event-study analysis and the panel regression analysis.

3. We explicitly write a regression model which is motivated by the SIRD model, and we clearly state the unconfoundedness assumption under which we may interpret our panel regression result as causal.

4. Given the comment of the referee that "the evidence regarding colleges is substantially less strong than the evidence for K-12 schools" as well as the space constraint after adding the event-study analysis, we decided to drop our focus on the effect of college openings and mainly focus on the effect of K-12 school openings in our revised manuscript. In particular, we made the following changes.

    - We moved Figure 2 in the previous manuscript and its related discussion on the effect of college openings on cases and deaths to SI Appendix.
    - We moved Table 2(b) in the previous manuscript on the effect of opening colleges and K-12 schools on the visits to restaurants and bars to SI Appendix because Table 2(b) was presented to highlight the effect of college openings on an increase in the visits to bars.
    - The paper's title has been changed to "The Association of Opening K-12 Schools with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis" by dropping "and Colleges" from the previous title.

5. For the death growth regression, we now use the outcome variable defined by the log difference over 21 days in weekly deaths rather than the log difference over 7 days because the available evidence from the CDC shows that the time lag between infection and death reporting is stochastic and spreads over at least 2 weeks (c.f., Table 2 of `https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html`). We also use the explanatory variables that are lagged by 35 days to capture the time lag of infection and death reporting, where we provide robustness checks with respect to the choice of lags in sensitivity analysis.

6. We now use the standard fixed effects estimator without bias correction rather than debiased fixed effects estimator for the effect of school openings on full-time workplace visits, staying home devices, visits to restaurants and bars in Tables 3 and Table S4 in the revised manuscript (which corresponds to Table 2(a)(b) in the previous manuscript). This is because the specification for mobility outcome does not have lagged dependent variables in the covariates and there is no need for implementing bias-corrections. The results are similar after changing from the debiased fixed effects estimator to the standard fixed effects estimator.

Reply to Referee 2's Comments

Comments

1. *"As always with observational analyses, the chief concern is confounding variables. There are many factors, such as other local policies and human behaviors, that may both affect covid case rates and be correlated with the decision of whether to re-open schools. The authors do a good job trying to account for many possible confounds, but they of course cannot control for everything."*

   Thank you for your positive comment. We agree that we "cannot control for everything" and we state the following in the subsection entitled "Limitations":

   > First, our study is observational and therefore should be interpreted with great caution. It only has a causal interpretation in a structural model under exogeneity assumptions that might not hold in reality (see the Model and Method in SI Appendix). While we present sensitivity analysis with a variety of controls including county dummies and interactions of state dummies and week dummies, the decisions to open K-12 schools and colleges/universities may be endogenous and correlated with other unobserved time-varying county-level factors that affect the spread of COVID-19. For example, people's attitudes toward social distancing, hand-washing, and mask-wearing may change over time (which we are not able to observe in the data) and their changes may be correlated with

school opening decisions beyond the controls we added to our regression specifications.

2. *"Figure 1d) shows a sharp increase in visits to full-time workplace visits around the time that schools open, with larger increases for In Person and Hybrid schools. The authors hypothesize that this may be a possible mechanism by which school openings increase cases, by getting adults back into work. However, an alternative story would be that there are confounding factors that are correlated with school openings and increase visits to workplace. There are a few pieces of evidence that are not entirely consistent with the authors' story of this being the mechanism and point more towards confounding:"*

   (a) *"The visits to workplaces seem to start to increase before schools open. If that is the case, then it seems unlikely this is caused by school openings. On the other hand, it's possible that this is due to measurement error about when schools open, in which case this should be noted."*

   This observed pattern of an apparent increase in the workplace visits before schools open is due to the way we pre-processed the variable for school opening dates as well as the visits to workplaces.

   First, the original data for school opening date is at the school district level, and different school districts opened their schools at different dates within the same county in many counties. Our county-level school opening date is constructed by taking the weighted average of school opening dates with the same teaching method across different school districts with student enrollment weights. As a result, some school districts open their schools before our constructed county-level school opening date, possibly contributing the increase in the visits to workplaces before the county-level school opening date. Second, our measure of the visits to workplace is based on the seven-day moving average, therefore, by construction, the visits-to-workplace variable is smoothed over 7 day. Because we roughly take the middle of 7-days moving average as the corresponding date to "Days since School Opening," by construction, there will be an increase in the visits to workplace 4 days before the school opening date if there is an increase in visits to workplaces on the school opening date.

   To show that the way we pre-processed our data is responsible for the apparent increase in the visits to workplaces before the school opening date, we take the subsample of counties in which all school districts share the same school opening date and construct the figures corresponding to Figure 1(c)(d) using the raw daily data for the visits to workplaces per device without taking 7-day averages.

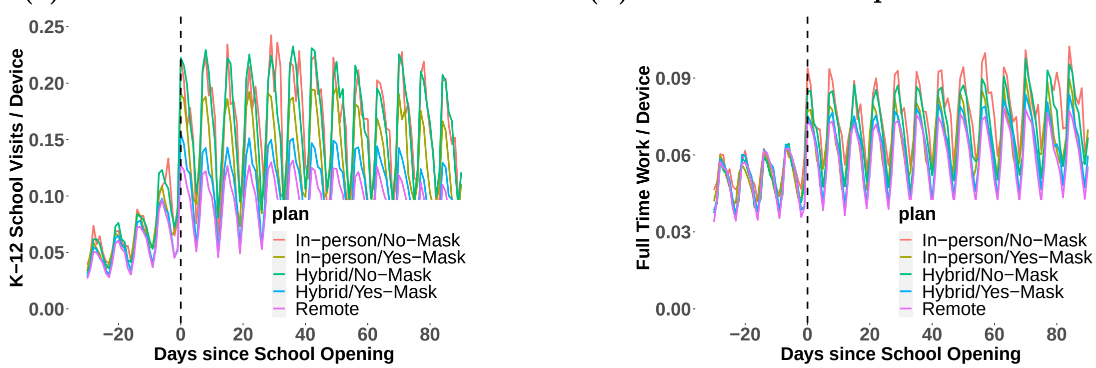
   Figure R6(d) shows that the average patterns of the visits to workplaces before the school opening are identical across different teaching modes, suggesting there is no confounding factor. Note that the sample size becomes substantially smaller than the original sample size (roughly one-half of the original sample size) because many counties have multiple school districts with different opening dates and therefore dropped out of the sample for Figure R6.

Following your suggestion, we present Figure R6(c)(d) in SI Appendix, Fig. S4 and added the following discussion in the footnote:

> Although the workplace visits appear to start increasing before school opens in Fig. R3(d), this reflects a measurement error about school opening dates as well as the use of the 7-day average visits. SI Appendix, Fig. S6(d) shows that a sharp increase in the daily visits to workplaces only happens on the day of school openings without any increase before for a subset of counties such that the school opening date is the same across all school districts within a county.

Figure R6: The evolution of visits to K-12 schools and full-time workplaces before and after the opening of K-12 schools using the daily frequency data for counties with the same school opening dates for all school districts within a county

(c) K-12 School Visits **(d) Full-Time Workplace Visits**



Notes: (c) and (d) plot the evolution of per-device visits to K-12 schools and full-time workplaces, respectively, against the days since K-12 school opening. Here, we restrict our sample to a subset of counties for which the school opening date is the same across all school districts within a county and the measurement of school visits and workplace visits is a daily measure rather than 7-days moving average.

(b) *"Visits to workplaces increase more for locations where schools are open without mask mandates than in places where they are open with mask mandates. Why should the mask policy in schools affect whether parents return to work or not?"*

Staff mask mandates are correlated with other mitigation measures at schools, including the availability of online instructions, as shown in Table R5. In particular, the correlation between no online instruction with no mask mandate for staffs is 0.13 (P-value<0.01) although there are many missing values for a variable for online instructions in our data set. One possible explanation for visits to workplaces increase more for locations where schools are open without mask mandates than in places where they are open with mask mandate is that the availability of online instruction allows some parents to choose not to send their children to schools and parents stay at home in such a case. This is consistent with the observed pattern in the visits to K-12 schools before and after the school opening as shown in Figure 1(c) in the manuscript as well as Figure R6(c) above, where the per-device visits to K-12 schools increases more for counties with mask mandates than for those without mask mandates even within the same mode of school opening (i.e., within "In-person" or "Hybrid").

To investigate this issue further, Figure R7 presents the evolution of visits to workplace and K-12 school over time by the availability of online instructions and compare them with the similar figure classified by mask mandates for staffs. As shown, visits to workplaces and K-12 school increase more for locations where schools are open without online instructions than in places where they are open with online instructions.

Following your comment, we emphasize the presence of high correlations across different mitigation measures together with a footnote that explains our rational for using "mask requirements for staff" as our main mitigation measure variable as follows:

> The measure of mask requirements for staff is highly correlated with other mitigation measures including mask requirements for students, prohibiting sports activities, and online instruction increases as shown in SI-Appendix, Table S3.[1]

We also cautioned an interpretation of the results regarding the estimated association of mask mandates for staffs with case growth as follows:

> These estimates likely reflect not only the effect of mask-wearing requirements for staff but also that of other mitigation measures. For example, school districts with staff mask-wearing requirements frequently require students to wear masks and often increase online instructions.

[1]MCH Strategic Data provides the school district level data on whether each school district adopts the following mitigation strategies: (i) mask requirements for staff, (ii) mask requirements for students, (iii) prohibiting sports activities, and (iv) online instruction increases, among other measures. We decided to use mask requirements for staff as the main variable for school mitigation strategy because it has a relatively smaller number of missing values. For panel regression analysis with the mask requirement variable, we drop counties from the sample when more than 50 percent of students in a county attend school districts of which mask requirements for staff is unknown or pending. Similarly, for specification with different teaching methods, we drop counties from the sample when more than 50 percent of students in a county attend school districts of which teaching methods are unknown or pending.

Table R5: Correlation across mitigation variables

	No mask for staffs	No mask for students	Yes sports for students	No online instruction
No mask for staffs	1.00			
No mask for students	0.74***	1.00		
Yes sports for students	0.14***	0.07***	1.00	
No online instruction	0.13***	0.12***	0.06**	1.00

Notes: Based on cross-sectional observations on October 1, 2020. **p<0.05; ***p<0.01

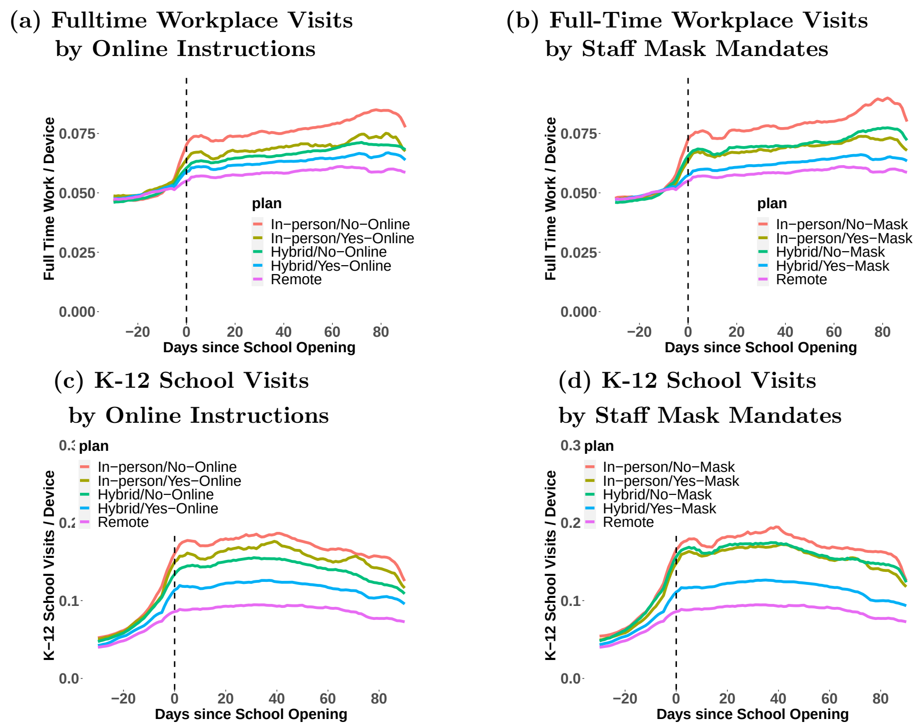
3. *"In my view, the evidence regarding colleges is substantially less strong than the evidence for K-12 schools."*

We agree that the evidence regarding colleges is less strong than the evidence for K-12 schools given that the estimated coefficient of college visits is sensitive to specifications. The estimated coefficient of college visits is less than one-third of that of K-12 school visits in column (1) of Table 2 while the median value of the per-device visits to college in October of 2021 is close to zero, suggesting that college openings may play a very limited role for spreading the COVID-19 relative to K-12 school openings perhaps except for college towns like Madison, WI or State College, PA. We state that our results suggest the limited role of college openings relative to K-12 school openings for virus transmissions in the main text.Given that our revised manuscript becomes longer after adding the new results on the event-study analysis, we decided to drop our focus on the effect of college openings and solely focus on the effect of K-12 school openings in our revised manuscript. In particular, we made the following changes.

- We moved Figure 2 and its related discussions on the effect of college openings on cases and deaths to SI Appendix. Our baseline specification for panel data regression still includes the per-device visits to colleges as a control.
- We also dropped Table 2(b) on the effect of opening colleges and K-12 schools on the visits to restaurants and bars because Table 2(b) was presented to highlight the effect of college openings on an increase in the visits to bars.
- The paper's title has been changed to "The Association of Opening K-12 Schools with the Spread of COVID-19 in the United States: County-Level Panel Data Analysis" by dropping "and Colleges" from the previous title.

(a) *"Figure 2 shows that case rates increase in Dane County, WI (home to University of WI) around the time that the university opens, with the largest increases among*

Figure R7: The evolution of visits to fulltime workplaces and K-12 schools before and after the opening of K-12 schools by mask mandates for staffs and by online instructions



college-age people. However, presumably the population that is college age also increased substantially when the school opened, so even if the case rate was constant, wouldn't we expect an increase in overall case numbers? Additionally, did the University require students to take tests upon returning to campus? This would also potentially increase the case counts."

We agree that college-age people increased substantially when the school opened and the increase in college-age population associated with the university opening contributes to an increase in the number of cases even if the case rate was constant. However, an increase in the number of cases from September to October of 2020 among college-age people is so large that the increase in college-age population associated with the university opening alone won't be able to explain the magnitude of case increases.

Specifically, the total population in Dane County, WI in 2020 is about 550,000 while the total enrollment at the University of Wisconsin-Madison is about 45,000. Therefore, the number of college-age people increased by the university opening is less than 10 percent of the total population in Dane County. Nearly 1,000 positive cases were confirmed on the UW-Madison campus by September 9, 2020, accounting for at least 74 percent of confirmed cases from September 1 to 8, 2020 in Dane county. This disproportionally large number of cases at the UW-Madison relative to the total number of population only makes sense if the case rates are higher among students at the UW-Madison than general population in Dane county.

On the other hand, we checked the testing policy at the UW-Madison and found that the UW-Madison offered no-cost testing to all students and the tests are mandatory upon returning to campus for those who returned to residence halls (source 1 and source 2). Therefore, we believe that an increase in confirmed cases among college-age people in Dane County must be partly driven by an increase in the number of tests. Unfortunately, we don't have county-level data on the number of tests and, therefore, it is not possible for us to analyze how much of an increase in confirmed cases is due to an increase in actual infections as opposed to an increase in the number of tests.

Our inability to separate the effect of an increase in tests from the effect of transmissions among university students is one of the reasons why we drop our focus on the effect of college openings in the revised manuscript. Given your comment, Figure 2 and its related discussion are now moved to SI Appendix and we point out that an increase in confirmed cases at the UW-Madison is partly driven by an increase in the number of tests.

(b) *"Table 2 columns (3) and (4) show that the estimated effect on college on college is 0.132 (**) in the base specification, but only 0.01 (not significant) when including interactions of K-12 In person visits with K-12 mask policies. Why should K-12 mask policies have such a large effect on the effect of college visits? Figure 3 likewise shows that several alternative specifications yield smaller, and sometimes insignificant, estimates for college."*

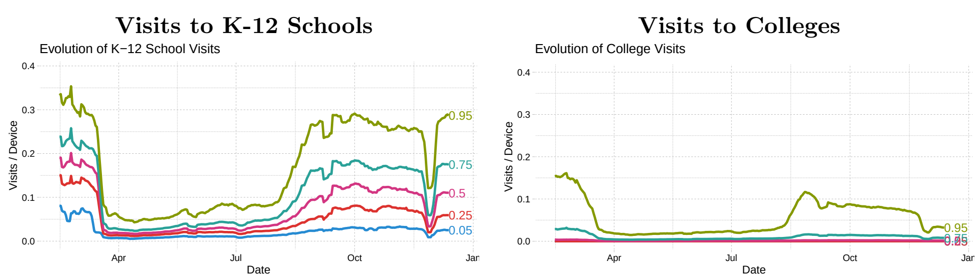
We agree that the estimated effects of college visits on case growth are small and sensitive to specifications. Following your comments, we drop our focus on the effect of college openings and discussed that the estimated coefficients on college visits are much smaller than that on K-12 visits and are sensitive to specifications in the main text as follows:

> The estimated coefficient of per-device visits to colleges is 0.14 (SE = 0.07) in column (1) of Table 2. With the change in top 5 percentile values of college visits between June and September as a benchmark for full openings, which is about 0.1, fully opening colleges may be associated with ($0.14\times0.1=$) 1.4 percentage points increase in the growth rates of cases. Therefore, the estimated association of opening colleges with case growth is much smaller than that of opening K-12 schools. Furthermore, alternative specifications

in columns (2) and (4) of Table 2 and those in Fig. 5 yield smaller, and sometimes insignificant, estimates for college visits. This sensitivity could be due to the limited variation in changes in college visits over time across counties in the data, where the 75 percentile value of college visits is consistently very small (SI Appendix, Fig. S3(f)). SI Appendix presents more discussions on the association of opening colleges with the spread of COVID-19, where Fig. S2 and S3 provide descriptive evidence that opening colleges and universities may be associated with the spread of COVID-19 in counties where a large public universities are located.

One reason for the small, sensitive estimates on college visits is that a majority of counties do not have any college and universities. Figure R8 shows that a median value of college visits is zero while even the 75 percentile value of college visits is consistently very small over time. Because the variation in changes in college visits over time across counties is very limited in the data, the association of college visits with case growth is much more difficult to identify than that of K-12 school visits.

Figure R8: Evolution of Visits to K-12 Schools and Colleges across U.S. counties



Notes: the figures report the evolution of various percentiles of corresponding variables in the title over time.

4. *"A concern with the dynamic panel regression that the authors use is measurement error. Since lagged case rates are included on the RHS, and the LHS variable is the change in case rates, measurement error in case rates will mechanically induce bias in the estimates. It would be useful to have some discussion of measurement error, and if possible, a discussion of the likely direction of this bias."*

We believe that this comment stems from the confusion we caused by not writing out the complete specification. To minimize the confusion about the exact definition of the outcome variable and the RHS variables, we now explicitly write our baseline regression specification in equation [5] followed by the definition of the variables as

$$\log Case_{it} - \log Case_{i,t-7} = \beta' \, Visit_{i,t-14} + \sum_{\tau=14,21,28} \beta_{y,\tau} \log Case_{i,t-\tau} \\ + \gamma' NPI_{i,t-14} + \theta \, Test_{it} + \alpha_i + \delta_{s(i),w(t)} + \epsilon_{it}, \tag{6}$$

where the outcome variable $\log Case_{it} - \log Case_{i,t-7}$ is the log-difference over 7 days in reported weekly cases with $Case_{it}$ denoting the number of confirmed cases from day $t-6$ to t.

The log of past cases on the LHS is lagged by 14 days while the RHS variable is the log difference between weekly cases at t and those at $t-7$. Thus, there is no overlap of the same variables between the LHS and the RHS, and measurement error in case rates will not mechanically induce bias in the estimates.

Furthermore, we *intend to* include the log of past weekly reported cases rather than that of the actual, but unknown, number of infected individuals in the RHS of (6) to capture people's voluntarily behavioral response to new information of transmission risks. It is the reported number of cases rather than the actual number of infected individuals that matters for people's behavioral response to the extent that the actual number of infected individuals are unknown beyond the reported cases. From this viewpoint, there is no "measurement error" on the RHS variable of the log of past weekly confirmed cases because that's what people are informed about. For clarification, we add the following discussion after presenting the specification (6) in the main text:

> Finally, the logarithm of past weekly confirmed cases denoted by $\log Case_{i,t-\tau}$ for $\tau = 14, 21$, and 28 are included in [4] to capture people's voluntarily behavioral response to new information of transmission risks. Controlling for past confirmed cases is important because people may voluntarily change their risk-taking behavior in response to the new information provided by the *confirmed* cases rather than the *actual*, but unknown, number of infected individuals.

5. *"I like how the functional form used by the authors is motivated by an SIRD model. However, it would be useful to have some robustness checks to understand how important the choice of functional form is to these results."*

Following your comments as well as the suggestions by the editor and the other referee, we now present event-study regression with both cases per 1000 and the log of weekly cases as dependent variables, providing robustness checks as reported in Fig. 2 and 3. In addition, we present the estimates of panel regression using the log of weekly cases—rather than the log difference—as dependent variable with and without lagged dependent variables on the RHS in SI Appendix, Tables S9-S10. The results from these robustness checks are consistent with the findings from our baseline panel regression analysis.

References

Blundell, Richard and Stephen Bond. 1998. "Initial conditions and moment restrictions in dynamic panel data models." Journal of Econometrics 87 (1):115–143. URL https://www.sciencedirect.com/science/article/pii/S0304407698000098.

Callaway, Brantly and Pedro H.C. Sant'Anna. 2020. "Difference-in-Differences with multiple time periods." Journal of Econometrics URL https://www.sciencedirect.com/science/article/pii/S0304407620303948.

Chen, Shuowen, Victor Chernozhukov, and Iván Fernández-Val. 2019. "Mastering panel metrics: causal impact of democracy on growth." In AEA Papers and Proceedings, vol. 109. 77–82.

Engzell, Per, Arun Frey, and Mark D. Verhagen. 2021. "Learning loss due to school closures during the COVID-19 pandemic." Proceedings of the National Academy of Sciences 118 (17). URL https://www.pnas.org/content/118/17/e2022376118.

Ford, Tamsin, Ann John, and David Gunnell. 2021. "Mental health of children and young people during pandemic." BMJ 372. URL https://www.bmj.com/content/372/bmj.n614.

Gadermann, Anne C, Kimberly C Thomson, Chris G Richardson, Monique Gagné, Corey McAuliffe, Saima Hirani, and Emily Jenkins. 2021. "Examining the impacts of the COVID-19 pandemic on family mental health in Canada: findings from a national cross-sectional study." BMJ Open 11 (1). URL https://bmjopen.bmj.com/content/11/1/e042871.

Goldhaber-Fiebert, Jeremy D., David M. Studdert, and Michelle M. Mello. 2020. "School Reopenings and the Community During the COVID-19 Pandemic." JAMA Health Forum 1 (10):e201294–e201294. URL https://doi.org/10.1001/jamahealthforum.2020.1294.

Goodman-Bacon, Andrew. 2018. "Difference-in-Differences with Variation in Treatment Timing." NBER Working Papers 25018, National Bureau of Economic Research, Inc. URL https://ideas.repec.org/p/nbr/nberwo/25018.html.

Hernán, M.A. and J.M. Robins. 2020. Causal Inference: What If. Chapman & Hall/CRC.

Isphording, Ingo E., Marc Lipfert, and Nico Pestel. 2021. "Does re-opening schools contribute to the spread of SARS-CoV-2? Evidence from staggered summer breaks in Germany." Journal of Public Economics 198:104426. URL https://www.sciencedirect.com/science/article/pii/S0047272721000621.

Leidman, Eva, Lindsey M. Duca, John D. Omura, Krista Proia, James W. Stephens, and Erin K. Sauber-Schatz. 2021. "COVID-19 Trends Among Persons Aged 0–24 Years — United States, March 1–December 12, 2020." MMWR Morb Mortal Wkly Rep 70. URL http://dx.doi.org/10.15585/mmwr.mm7003e1.

Lessler, Justin, M. Kate Grabowski, Kyra H. Grantz, Elena Badillo-Goicoechea, C. Jessica E. Metcalf, Carly Lupton-Smith, Andrew S. Azman, and Elizabeth A. Stuart. 2021. "Household COVID-19 risk and in-person schooling." medRxiv URL https://www.medrxiv.org/content/early/2021/03/01/2021.02.27.21252597.

Parcha, Vibhu, Katherine S. Booker, Rajat Kalra, Seth Kuranz, Lorenzo Berra, Garima Arora, and Pankaj Arora. 2021. "A retrospective cohort study of 12,306 pediatric COVID-19 patients in the United States." Scientific Reports 11 (1):10231. URL https://doi.org/10.1038/s41598-021-89553-1.

Roth, Jonathan and Pedro H. C. Sant'Anna. 2021. "When Is Parallel Trends Sensitive to Functional Form?"

Sun, Liyang and Sarah Abraham. 2020. "Estimating dynamic treatment effects in event studies with heterogeneous treatment effects." Journal of Econometrics URL https://www.sciencedirect.com/science/article/pii/S030440762030378X.

Takaku, Reo and Izumi Yokoyama. 2021. "What the COVID-19 school closure left in its wake: Evidence from a regression discontinuity analysis in Japan." Journal of Public Economics 195:104364. URL https://www.sciencedirect.com/science/article/pii/S0047272720302280.

von Bismarck-Osten, Clara, Kirill Borusyak, and Uta Schönberg. 2020. "The Role of Schools in Transmission of the SARS-CoV-2 Virus: Quasi-Experimental Evidence from Germany." CReAM Discussion Paper Series 2022, Centre for Research and Analysis of Migration (CReAM), Department of Economics, University College London. URL https://EconPapers.repec.org/RePEc:crm:wpaper:2022.