Supplementary Appendix

Supplement to: Cowger TL, Murray EJ, Clarke J, et al. Lifting universal masking in schools — Covid-19 incidence among students and staff. N Engl J Med. DOI: 10.1056/NEJMoa2211029

This appendix has been provided by the authors to give readers additional information about the work.

Table of Contents

SUPPLEMENTARY METHODS & RESULTS

Inclusion/Exclusion Criteria

As summarized in **Table S1** below, in total, N=400 districts reported COVID-19 data to DESE. After excluding n=211 charter, vocational, and technical school districts and restricting to school districts within the Boston-Newton-Cambridge New England City and Town Area (NECTA), n=79 schools remained (See **Fig. S1**). We additionally excluded n=7 school districts with unreliable data if they met the following criteria:

1. >10 weeks with zero reported cases among students and staff
   OR
2. More than 5 reporting weeks with either:
   a. Zero reported cases among students and staff and at least 1 positive pool test
      OR
   b. Zero cases reported among students and staff and >10 cases reported on average over the past 4 weeks

After these n=7 exclusions, our final sample size of n=72 school districts (**Table S1).** One school district in our study, Brookline Public Schools, reinstated masking requirements on May 23, 2022. For this school district, we excluded study weeks after masking was reintroduced (i.e., the final 3 weeks of the study period).

Case Ascertainment and Routine Reporting

Throughout the 2021-2022 school year, the Massachusetts Department of Secondary and Elementary Education (DESE) required that all public school districts, charter school districts, approved special education schools, and education collaboratives report COVID-19 cases among students and staff to DESE on a weekly basis. Students and staff who test positive for COVID-19 outside of school were instructed to notify school staff and follow DESE's isolation guidance. School district staff then reported the number of cases among students and staff through a Security Portal application by 5pm Wednesday of each reporting week; any school/district that does not report their case count by this deadline was be assumed to have zero cases during the reporting period (correction for this is further discussed above in the Inclusion/Exclusion criteria section and in **Table S3** below). DESE releases a weekly report with school district-level counts of COVID-19 cases among students and staff (data used in this study), school district-level pooled testing results and a statewide summary of testing across school districts.

In addition to required case reporting, DESE also strongly recommended and provided full funding for school districts to opt-in to one or more of their state-sponsored testing programs. During 2021-2022 school year, DESE offered four different school testing programs – in-school symptomatic testing, routine pooled testing, "Test-and-Stay"/close contact testing, and weekly take-home rapid testing (further described in **Table S2**). For schools participating in these programs, DESE provided training, staffing, supplies, and operational and logistic support at no cost to school districts, and approximately 95% of all Massachusetts school districts participated in at least one of these programs.

Both symptomatic testing and routine pooled testing were supported by DESE throughout the entirety of the 2021-2022 school year. However, starting in January, DESE strongly recommended replacement of “Test-and-Stay”/close contact surveillance testing with free weekly take-home rapid antigen to participating schools (**Table S2**). DESE does not publicly release information about which testing programs school districts selected or participation in these programs by school district, limiting our ability to examine whether testing practices differed between districts and/or over time. However, a strength of our difference-in-differences design is that neither differences in testing practices between districts that do not vary over time (e.g., differential participation in routine pooled testing programs) nor changes in testing programs over time that are consistent across intervention and control districts (e.g., switch from test-and-stay/close contact testing to take-home rapid testing) can introduce bias. We discuss and assess the potential bias introduced by differences in testing programs below in the Supplementary Appendix section ‘*Robustness of Results to Differences in Testing Programs*,’ below.

Differences-in-Differences Analysis Methodological Details

Model Details

For our main analysis, we conducted a difference-in-differences (DiD) analysis with staggered implementation to compare weekly incidence of COVID-19 in school districts that lifted mask requirements to incidence in school districts where masking requirements had not yet been lifted. For this analysis, we utilized the approach introduced in Callaway and Sant’Anna 2021,[1,2] for differences-in-differences with multiple time periods and variation in treatment timing. This approach also allows for the intervention effect to vary over time (e.g., impact of masking may differ depending on duration of implementation or with changing community rates of COVID-19).

Using the `did` R package from Callaway & Sant’Anna,[1,2] we first estimated the group-time Average Treatment Effect among the treated, $ATT(g,t)$, at each week post-intervention for each intervention group using not-yet-treated school districts as controls fit using multivariable regression with a linear link. When adjusting for covariates, we report the marginalized ATT estimates, which do not require effect homogeneity assumptions. Additional details can be found in Callaway and Sant’Anna 2021,[1,2] however briefly, the `did` package estimates group-time average treatment effects($ATT(g,t)$) for each time period, t, by first subsetting to observations in timepoints t and $g-1$ and to observations in group g (i.e., $G_g = 1$) or groups not-yet-treated by time t (i.e., control districts for that time point) and fits the population linear regression:

$$Y = \alpha_1^{g,t} + \alpha_2^{g,t} \cdot G_g + \alpha_3^{g,t} \cdot I\{T = t\} + \beta^{g,t} \cdot \left(G_g \times I\{T = t\}\right) + \varepsilon^{g,t}$$

- Y is the weekly COVID-19 case rate per 100,000 students and/or staff
- G_g Is an indicator variable for whether a district first lifts their masking requirement in week g[1]
- $T = t$ is an indicator for the current modelled week

In the above equation, $ATT(g,t) = \beta^{g,t}$. This approach effectively creates school district- and time-fixed effects by utilizing within-school district change and estimating each individual time point separately.

Next, again using the `did` package, we employed two different aggregation schemes to summarize these group-time average treatment effects – event study/dynamic effects aggregation and calendar

[1] In our study, we have three such timepoints, study week 26 (03-Mar-2022, n=46 school districts); study week 27 (10-Mar-2022, n=17 school districts); and study week 28 (17-March-2022, n=7).

time aggregation (see Callaway and Sant'Anna 2021).[1] Event study aggregation uses a weighted-average of $ATT(g,t)$ estimates grouped by length of time since lifting of masking requirements. The results from this event study/dynamic aggregation are shown in **Figure 2** of the main text, where time $t = 0$ corresponds to the average effect of removing masking requirements in the first week masking requirements are lifted, and negative numbers indicate weeks prior to lifting requirements). We also calculated calendar time aggregation group $ATT(g,t)$ estimates in each calendar week, which allowed us to estimate the weekly and cumulative effect of lifting masking requirements over the 15 calendar weeks post-lifting of the statewide requirement, which we present in **Table 1** of our main text. Consistent with Callaway and Sant'Anna 2021,[1] we first calculated a weekly measure as a weighted average across all groups that have lifted their masking requirements by time, t, as

$$\theta_c(t) = \sum_{g \in \mathcal{G}} I\{t \geq g\}\, P(G = g\, |G \leq t)\, ATT(g,t)$$

We then extend this parameter to calculate the cumulative effect of lifting masking requirements from the time the statewide requirement was lifted ($t = 26$) up to time t, for all school districts that have lifted their masking requirement by that period, as reported in **Table 1** of our main text:

$$\theta_c^{cumu}(\tilde{t}) = \sum_{t=26}^{\tilde{t}} \theta_c(t)$$

We obtained standard errors and corresponding 95% confidence intervals for our estimates using the multiplier bootstrap procedures clustered at the school district-level as described in Callaway and Sant'Anna 2021.[1]

Assumptions and Test for Parallel Pre-Trends

An identifying assumption for our difference-in-differences analyses is that COVID-19 case rates for the school districts that first lifted their masking requirements in week g would have followed parallel paths to the school districts that had not yet lifted their masking requirements up to that time point, had these districts not lifted their masking requirements. To verify the plausibility of this assumption, we assessed whether trends were parallel in the pre-treatment period (i.e., in the period prior to removing masking requirements) by (1) reviewing the coefficients and standard errors from our event-study regression and (2) conducting a formal statistical test (Cramér-von Mises [CvM]) for parallel trends in the pre-period using the 'not-yet-treated' school districts as comparison units[3] following the methods outlined in Callaway & Sant'Anna, 2018.[4] Both approaches provide support for parallel trends holding in the pre-period.

In event study analyses the pre-period coefficients were largely centered around zero for staff and students combined and considered separately (**Fig. 2**, **Table S4**) – 24 of 26 (92%) pre-period coefficients were not significantly different from 0, and the remaining 2 of 26 (8%) coefficients did significantly differ from 0. Note that this does not adjust for multiple testing. These results are consistent with Cramér-von Mises tests (p>0.05) for staff and students combined and considered separately, suggesting no clear evidence of violations of the parallel trends assumption in the pre-period.

Limited Treatment Anticipation Assumption

In addition to parallel trends assumption, identification of causal effects requires an assumption of limited treatment anticipation.[1] Under a strict no-anticipation assumption, we assume that the lifting of masking requirements can have no causal effect on COVID-19 rates prior to its implementation. In our study, it is possible that some changes to masking policies were announced and/or anticipated a week or two ahead of their implementation (e.g., some school districts may have been less strict about adhering to masking requirements after lifting of the statewide requirement). The no-anticipation assumption can be relaxed to a less strict limited anticipation assumption where we can accommodate anticipation of change in policies if the anticipation period is specified. In the case of our study, we compared the results of our main analysis (i.e., zero weeks anticipation) to anticipation periods of 1,2, and 3 weeks prior to lifting if masking requirements. As shown in **Figure S6**, our effect estimates among staff and students combined and separately were robust to changes in anticipation periods.

Adjustment for Time-Varying Covariates

DiD methods are not biased by unmeasured time-invariant confounders or unmeasured time-varying confounders with consistent trends across intervention and control groups, strengthening causal inference.[3,5,6] Instead, to introduce bias, confounders have to impact the outcome of interest and change over time differentially between groups. We adjusted for several time-varying covariates in sensitivity analyses, listed in part (E) of **Table S3**. Full results from these sensitivity analyses are shown in **Figure S5**, however, we also show the trends in several time-varying covariates by week masking requirements were lifted for school districts in **Figure S7.**

Sensitivity Analyses

We conducted several sensitivity analyses to assess whether our results were robust to various data cleaning steps and model specifications. Our sensitivity analyses fell into five broad categories: (A) data cleaning; (B) population weighting; (C) smoothing; (D) modified control groups; and (E) adjustment for time-varying covariates. For each of these categories, additional details of these analyses are described in **Table S3**, below. Results of the sensitivity analyses described below can be found in **Figure S5**.

Causal Inference Considerations for Adjusting for Background Community-Levels of COVID-19

A key question in studies of schools is how background community COVID-19 rates interact with school-specific rates, and the correct analytic decision on whether to include community levels as a confounder or not is not straightforward. In sensitivity analyses, we found benefits of universal masking requirements persisted even after controlling for several measures of community COVID-19 incidence (**Fig. S5**). We did not prioritize these results in our main analyses, however, because we believe it is more appropriate to consider community rates of COVID-19 as part of the causal effect of school masking policies rather than a source of bias (i.e., a mediator rather than a confounder). A growing body of evidence suggests schools as a driver of COVID-19 community burden.[7–9] School and community COVID-19 levels are so closely linked that it is difficult to rule out at least some of the variation in community COVID-19 rates being a direct consequence of changing school case rates.[63] In our own study, we noted similar trajectories between COVID-19 case rates in schools and measures of COVID-19 burden in the surrounding communities (**Fig. S8**). However, we also found that COVID-19 case rates in schools began to increase several days *before* increases in COVID-19 burden in the community, and that this effect was more pronounced (i.e., earlier increases) among school districts that lifted masking requirements compared to Boston and Chelsea where masking requirements were sustained (**Fig. S8**). This ordering of timing (i.e., increases in schools prior to communities) provides further rationale for our decision to consider community case rates a mediator rather than confounder in our main analyses. In addition, while some community-level policies changed during Spring 2022, these changes did not perfectly coincide with school mask mandates being rescinded, nor did the community mask policies always align with school mask policies.

A second reason to consider community COVID-19 cases is that there may be spillover between school districts or communities as individuals move between areas. This is not an issue of confounding, but rather a potential threat to the validity of estimating causal effects of COVID-19 prevention policies. Spillover between communities and school districts may reduce the difference in COVID-19 rates and has the potential to alter the case growth trajectory in masked schools. This latter issue is a potential threat to the parallel trends assumption required for difference-in-differences analyses. To address this, we used the staggered policy adoption model of difference-in-differences.[1,2,4,10] This approach uses all pre-policy change data as unexposed data, and aligns the comparisons for school districts which unmask to contemporaneous control districts, allowing us to estimate the impact of removing masks, uncoupled from calendar time. As a result, we anticipate that any remaining impact of spillover on our findings would be to decrease the estimated impact of removing masks, making our results an underestimate of the harms caused by removing mask mandates.

Robustness of Results to Differences in Testing Programs

In addition to sensitivity analyses described above, we also wanted to ensure that differences in COVID-19 case rates between masked and unmasked districts were not due to differences in testing practices or definitions between school districts. Throughout the school year, DESE provided guidance and full funding for standardized COVID-19 testing programs in school districts. Prior to January 2022, DESE strongly recommended that schools participate in three types of testing programs: in-school symptomatic testing, Test-and-Stay/close contact testing, and routine pooled testing.

As described above, our study relied on reported case rates, and because of this, it is likely that some COVID-19 cases went undetected/unreported, introducing measurement error into our outcome. If this under detection/underreporting of cases was not differential by school district and varying over time, (i.e., non-differential), it is possible that our estimates of removing masking requirements are underestimates of the true effect of lifting masking requirements in schools in our differences-in-differences analysis. However, if the under ascertainment/reporting varied over-time and between districts, the magnitude and direction of potential bias is not straightforward, and these scenarios are discussed in further detail in the subsequent paragraphs.

Across study districts, the number of pooled screening tests per student stayed relatively consistent across the 15 weeks post-lifting of the statewide masking requirement providing supporting evidence that changes in pooled testing programs likely did not contribute substantially to our results (**Fig. S9**). In addition, during the period post-lifting of masking requirements number of pooled test per student were highest in school districts that sustained masking requirements than in any other group of comparison school districts by week masking requirements were lifted for school districts (**Fig. S9**).

As part of the Test-and-Stay/contact tracing program, close contacts who were "…exposed to a COVID-19 positive individual in the classroom, so long as the individuals were spaced at least 3 feet apart, or on the bus, while both individuals were masked, were exempt from testing and quarantine response protocols."[11] Therefore, prior to the lifting of the statewide mandate when all school districts had mask mandates in place, testing of close contacts depended on whether both individuals were wearing masks.

However, the vast majority of school districts had discontinued Test-and-Stay (i.e., testing of unmasked close contacts) prior to the lifting of the statewide masking requirement, and were therefore no longer using this definition in their testing criteria when masking requirements were lifted. On January 18, prior to the lifting of the statewide masking requirement, DESE strongly recommended that states **discontinue test-and-stay/close contact testing** and transition to weekly at-home rapid antigen surveillance testing provided by DESE starting the week of January 31.

> *"Updated testing guidance from DESE and DPH recommend that districts select the new option within the statewide testing program to offer weekly rapid at-home antigen tests to students and staff and discontinue contact tracing and Test and Stay programs. As such, for districts selecting this new option, individuals will no longer be identified as close contacts by school health professionals." (Page 5, Jan 18 guidance)*[11]

In addition, DESE states that "…districts and schools that choose to maintain Test and Stay will continue contact tracing and will not be eligible to receive rapid antigen at-home tests."

Importantly, the definition exempting of masked close contacts from testing only occurred under the Test and Stay program, and testing under other programs, including at-home surveillance testing, did not depend on individual masking status. However, if a large number of unmasked school districts continued Test and Stay programs through periods where masks were not required, there is a potential to introduce bias since more testing would have been done in schools that lifted masking requirements compared to school districts that remained masked.

DESE does not report which individual school districts chose to continue Test and Stay/close contact testing, however, they do provide information on the total number of tests conducted and COVID-19 cases identified through Test and Stay programs statewide. Statewide participation in Test and Stay decreased dramatically following the DESE's January 18 updated testing guidance from an average of ~35,000 close contacts tests/week to ~3,500 close contacts tests/week (**Fig. S10**). Statewide, only n=820 total COVID-19 cases were identified through Test-and-Stay in all MA school districts (i.e., not only districts included in our study) over the 15 weeks post-lifting of the statewide masking order.

In the most extreme case, if we assume:

1) All districts that lifted masking did not follow DESE guidance and continued Test and Stay programs (i.e., increased testing of unmasked close contacts)
2) All COVID-19 cases identified statewide through Test and Stay occurred in unmasked study districts (i.e., if close contact testing was only continued in school districts included in our study)
3) The cases identified through Test and Stay/close contact testing would not have been identified under any other testing program (e.g., weekly at-home rapid testing)

Under these assumptions, the n=802 total cases identified through Test and Stay can represent at most only 6.9% (95% CI: 5.4%, 9.4%) of the 11,901 (95% CI: 8,651, 15,151) COVID-19 cases we estimate were attributable to lifting of masking requirements (**Table 1**).

In addition, it is unlikely the above extreme assumptions are met. Study districts that lifted masking requirements account for <30% of students and staff statewide, and all 10 of the top 10 largest control school districts in our study (~30% of staff/students) ended Test and Stay/close contact testing before the statewide mask mandate was lifted (**Table S6**).

In summary, even after the most extreme assumptions in accounting for potential differences in close contact testing for masked vs. unmasked districts, the impacts of removing masks remain substantial.

Calculation of School Days Missed

Weekly school absences were not publicly available at the state-level for school districts in our study. Instead, to estimate the number of missed school days attributable to lifting of mask requirements, we started with the estimated number of cases attributable to lifting of the masking requirements (presented in **Table 1** of main results). We multiplied the attributable cases by the minimum 5-day isolation requirement set by DESE during our study period. We then multiplied this product by the proportion of calendar days that children were in school to account for the fact that not all isolation days after a COVID-19 case is identified would fall on school days (i.e., school holidays, weekends, etc.). To get the proportion of calendar days during our study we used the 2021-2022 Boston Public Schools calendar[2] and determined that there were 181.5 school days out of 290 calendar days and therefore in total, 62.6% of calendar days of the school year are days students/staff spend in schools. The results of these calculations are shown in **Table S5**.

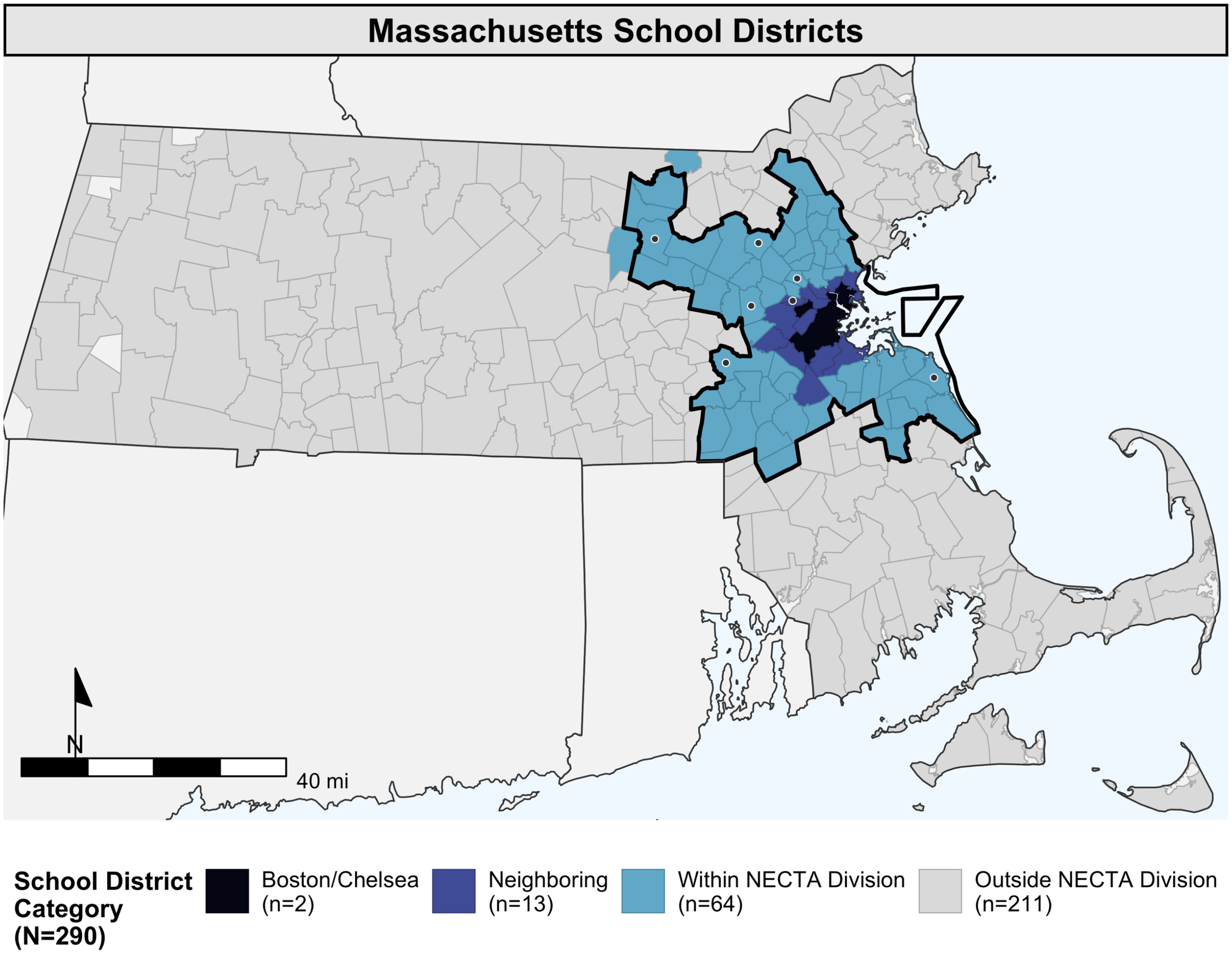
Finally, to reflect the statistical uncertainty in our estimates, we present the lower confidence interval for the estimated missed number of school days, as this represents a conservative estimate based on our models (e.g., "… this translated to a minimum of approximately 17,500 days of school absence in students and 6,500 days of staff absence over 15 weeks"). The numbers we present in our results section are highlighted and underlined in **Table S5**.

As mentioned above, school absences were not systematically reported across school districts in our study in DESE data. However, one school district in our study – Lexington, MA, a comparison district ~10 miles from Boston – did report weekly COVID-19 absences among students and staff throughout the 2021-2022 school year. For this district, we compared average weekly absences in weeks where masking was required to weeks where masking was optional. We found that average weekly staff and student COVID-19 absences were 50% higher in mask-optional weeks compared to prior weeks when masking was required (**Fig. S11**).

[2] 2021-2022 Boston Public Schools Calendar, available: https://www.bostonpublicschools.org/cms/lib/MA01906464/Centricity/Domain/4/BPS%20Cal%20SY22_FINAL.pdf

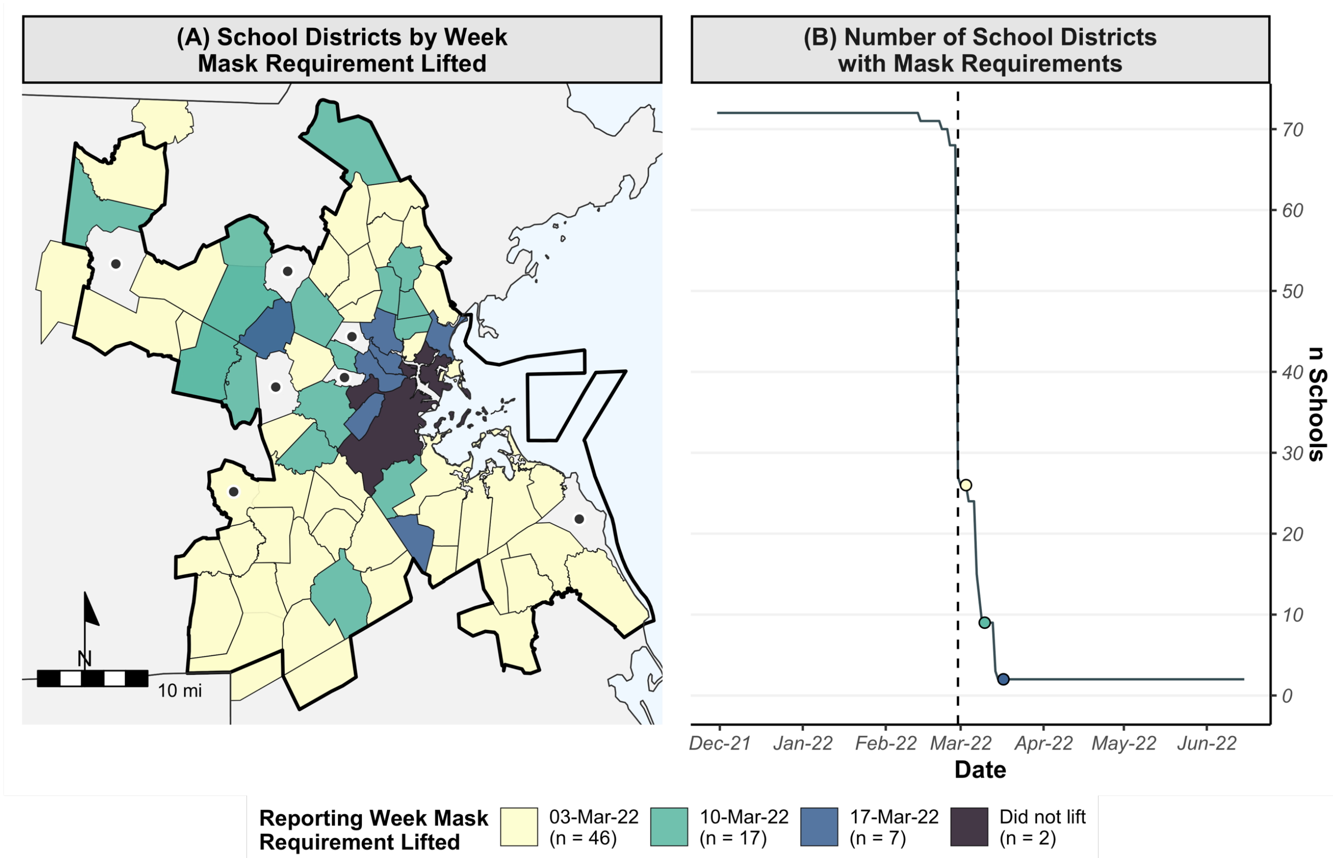
SUPPLEMENTARY FIGURES

Figure S1. *Map of Massachusetts school districts for inclusion in the present study. Black dots indicate the n=7 school districts[3] within the Boston-Cambridge-Newton NECTA division that were excluded from the study due to unreliable data.*



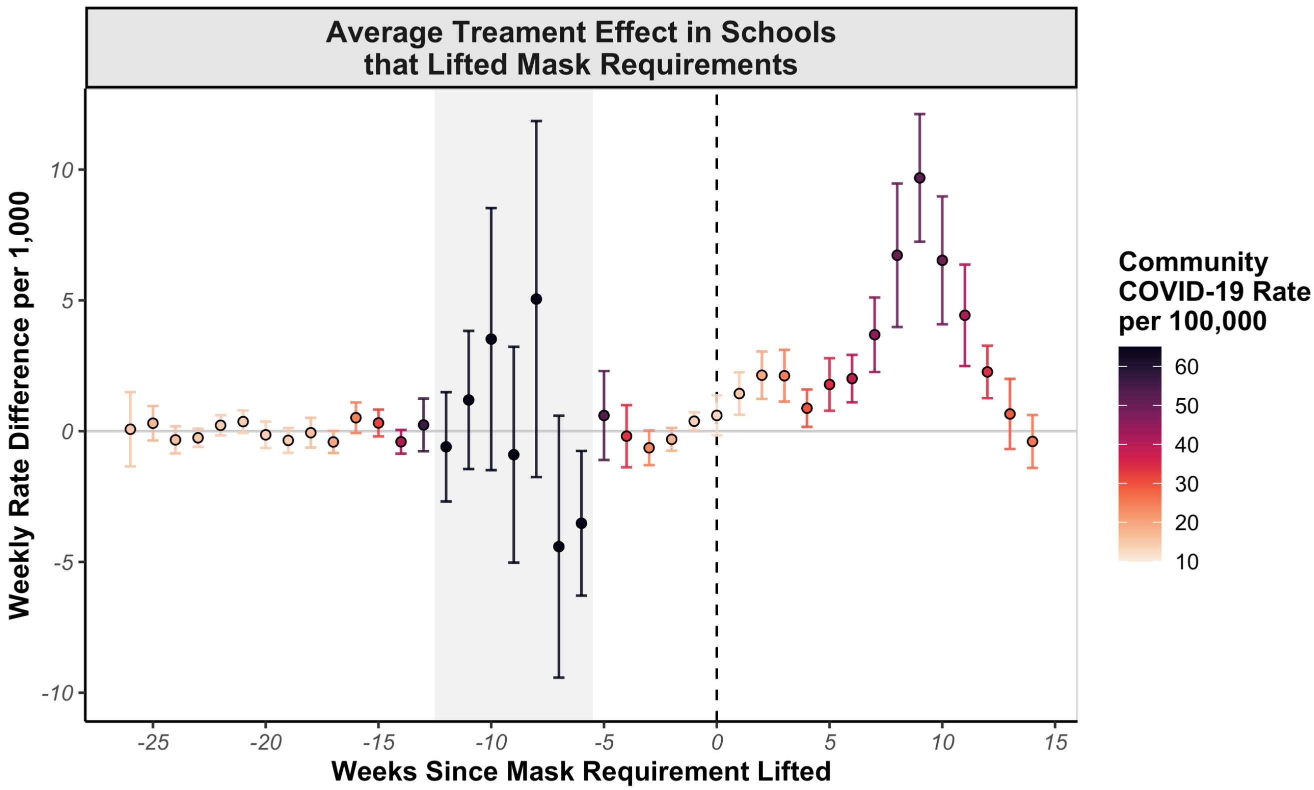
[3] Excluded school districts: Arlington, Bedford, Harvard, Scituate, Sherborn, Weston, and Watertown

Figure S2. *(A) Map of school districts in Boston-Cambridge-Newton NECTA division by reporting week in which mask requirements were removed and (B) number of school districts with mask requirements remaining in place over time*[4]



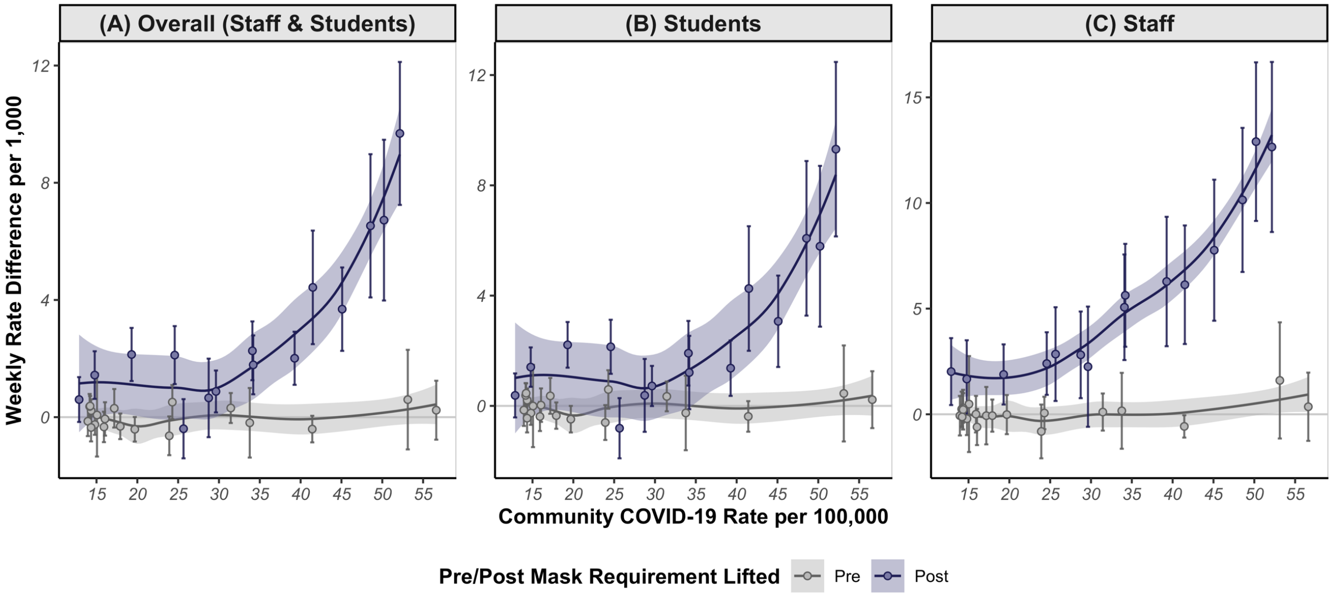
[4] Black dots denote the n=7 school districts excluded from the analysis due to unreliable data reporting, see **Figure S1**

Figure S3. *Weekly differences in rate of COVID-19 among staff and students in school districts that lifted masking requirements compared to school districts that had not yet lifted their masking requirements colored by background community COVID-19 rate*[5]



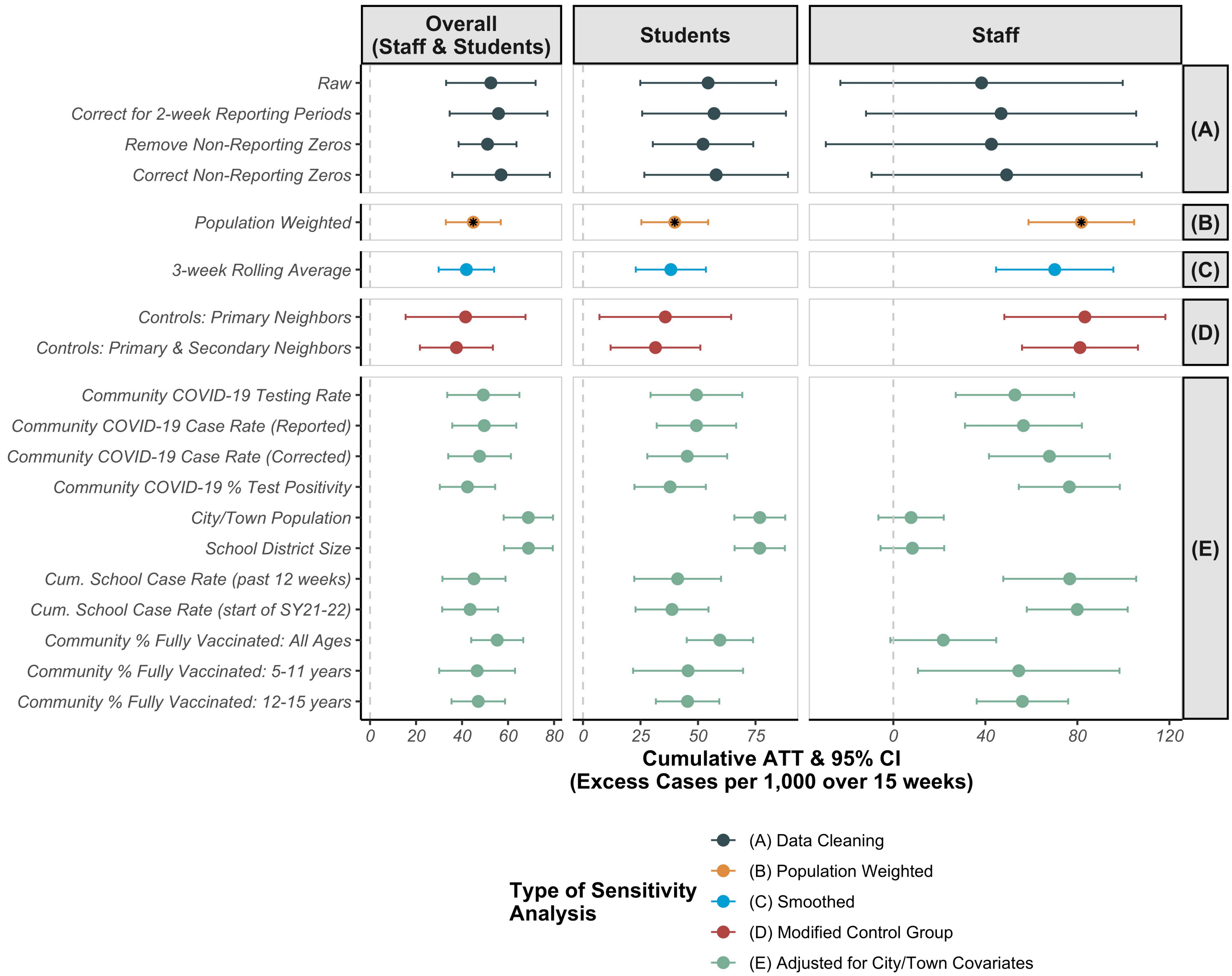
[5] Estimates are from difference in differences models and identical to those depicted in the top panel of **Figure 2A** in the main text; Darker colors indicate higher background COVID-19 rates in the corresponding city/town; Grey band on the plot depicts the initial BA.1 Omicron wave in December 2021-January 2022

Figure S4. *Relationship between COVID-19 case rates in surrounding communities (x-axis)[6] and weekly effect estimate sizes for the average treatment effect of removing masking requirements (y-axis) before (pre-period) and after (post-period) masking requirements were lifted*



[6] X-axis attenuated at a weekly COVID-19 community case rate of 55 per 100,000 persons to show the range community case rates observed both in pre- and post- masking period (i.e., excluding highest Omicron peak values in early January 2022)

Figure S5. *Cumulative effect estimates for the impact of lifting mask requirements overall and for students and staff under various sensitivity analyses[7] including (A) varied data cleaning and suppression steps, (B) weighted by school district size (i.e., population),[8] (C) smoothed using 3-week rolling averages, (D) using varied control school districts, and (E) adjusting for various covariates (See Supplementary Appendix and Table S3 for additional details)*



[7] These estimates correspond to estimates presented in **Table 1** of main text
[8] Black asterisk indicates the results presented in main analyses (i.e., results in main text are population weighted)

Figure S6. *Results of sensitivity analysis assessing impact of 1-3 weeks of treatment anticipation compared to effect estimates observed in main analysis (A) overall, (B) among students, and (C) among staff*

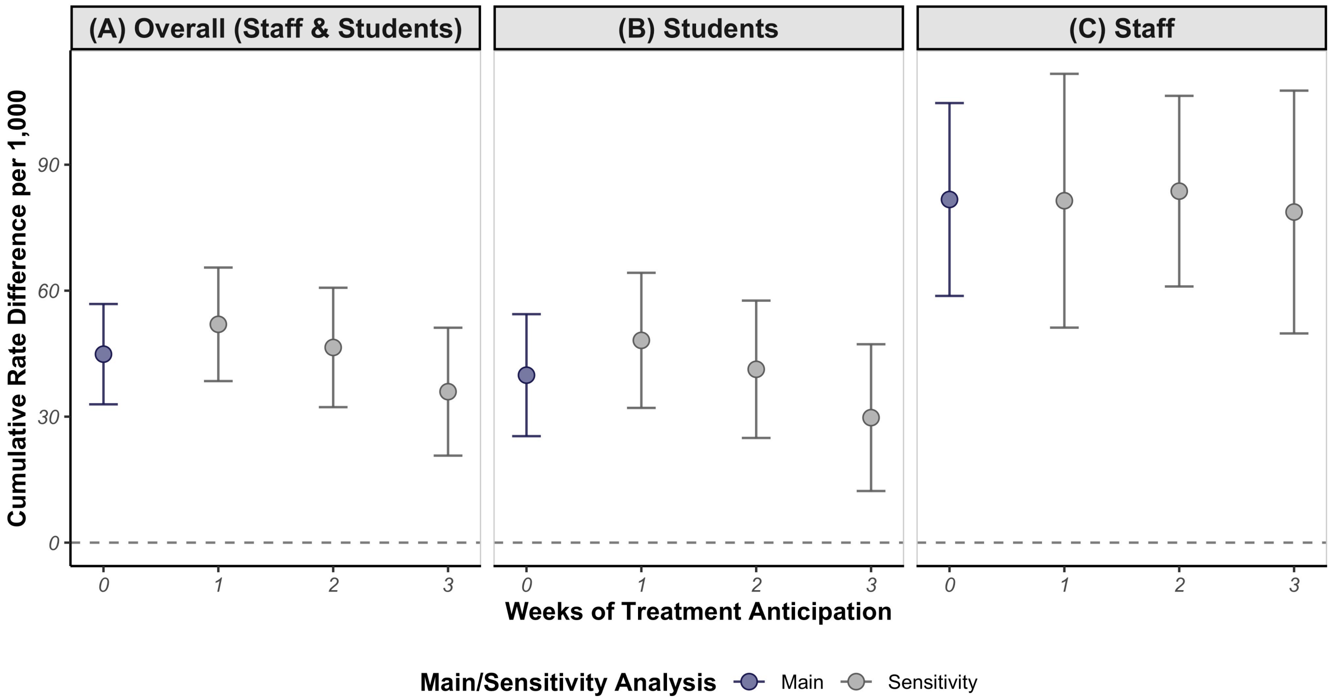
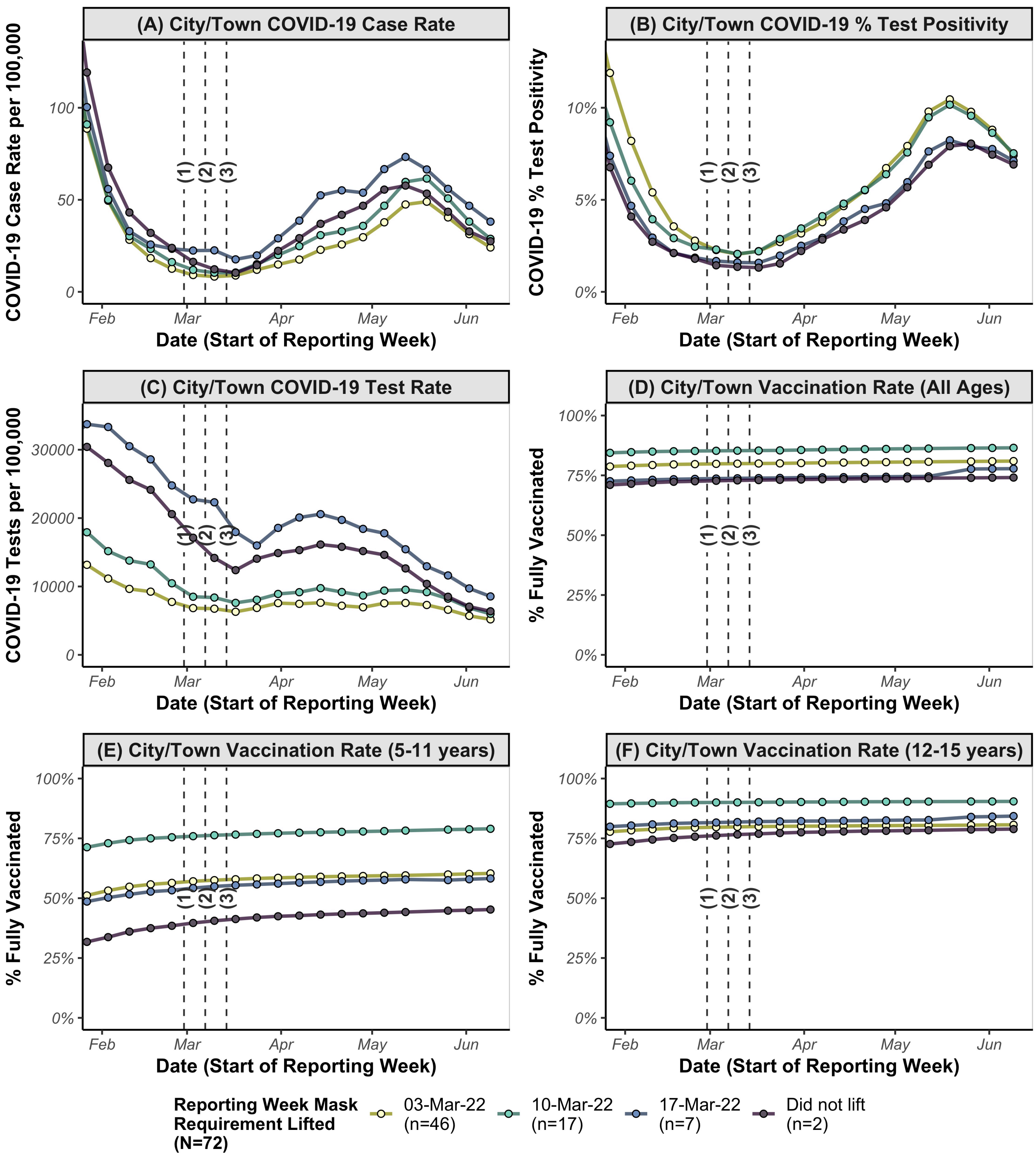


Figure S7. *Selected time-varying covariates by week masking requirements were lifted.*[9]

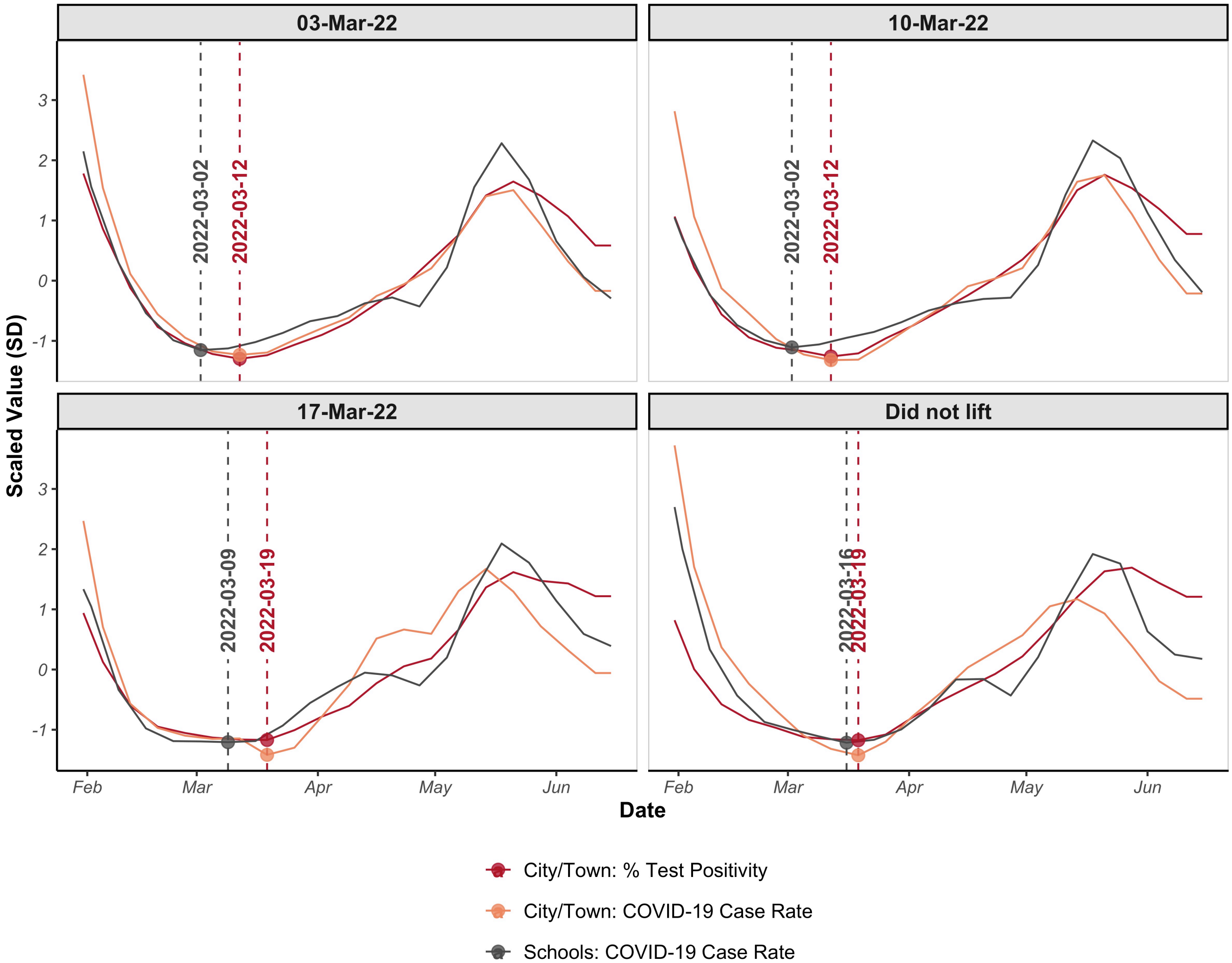


(1) Week of 28-Feb-22: DESE Statewide Mandate lifted (n = 46); (2) Week of 07-Mar-22 (n = 17); (3) Week of 14-Mar-22 (n = 7)

---

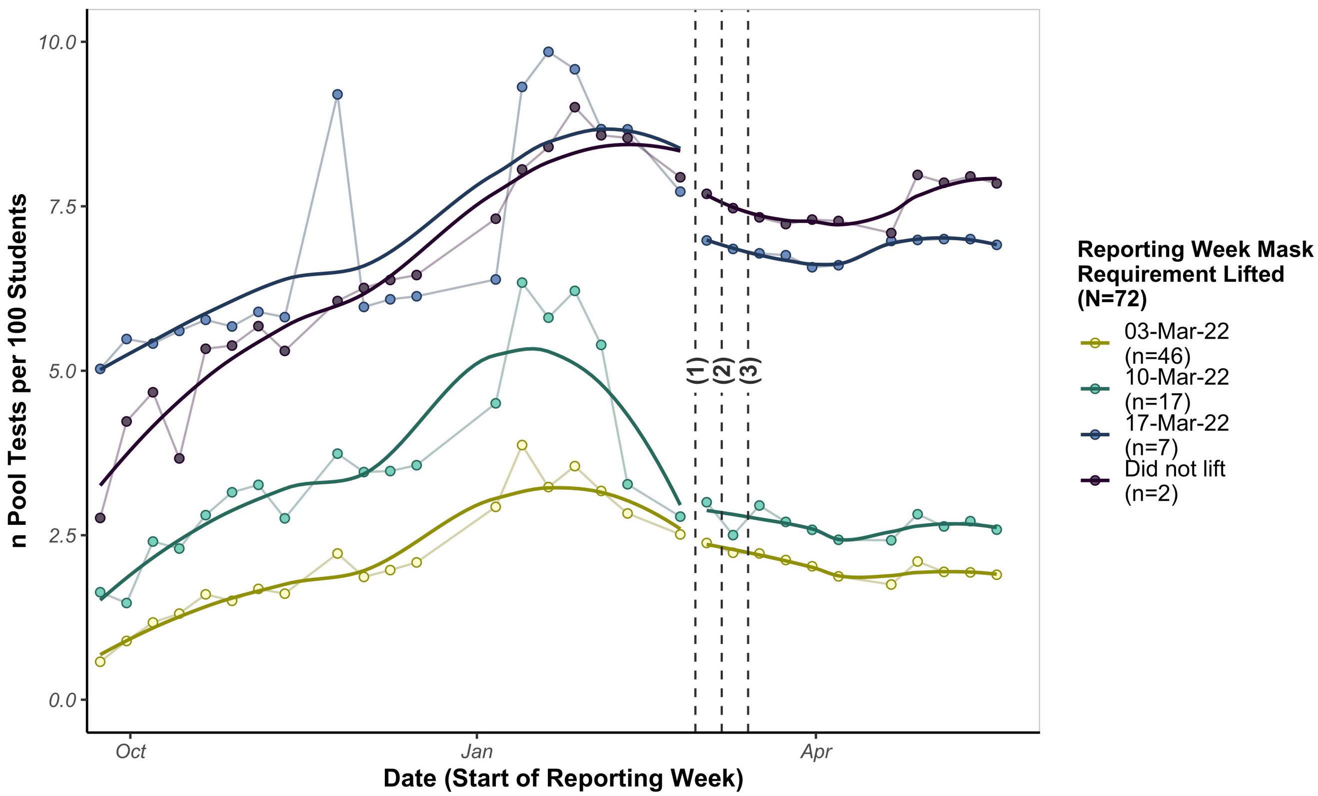
9 Dates on the x-axis were restricted to the period immediately before and after universal school masking requirements were lifted statewide and in most school districts. Difference-in-differences analysis includes all weeks in the 2021-2022 school year.

Figure S8. *COVID-19 case rates in schools and COVID-19 percent test positivity and case rates in surrounding cities/towns by week universal school masking requirement was lifted*[10]



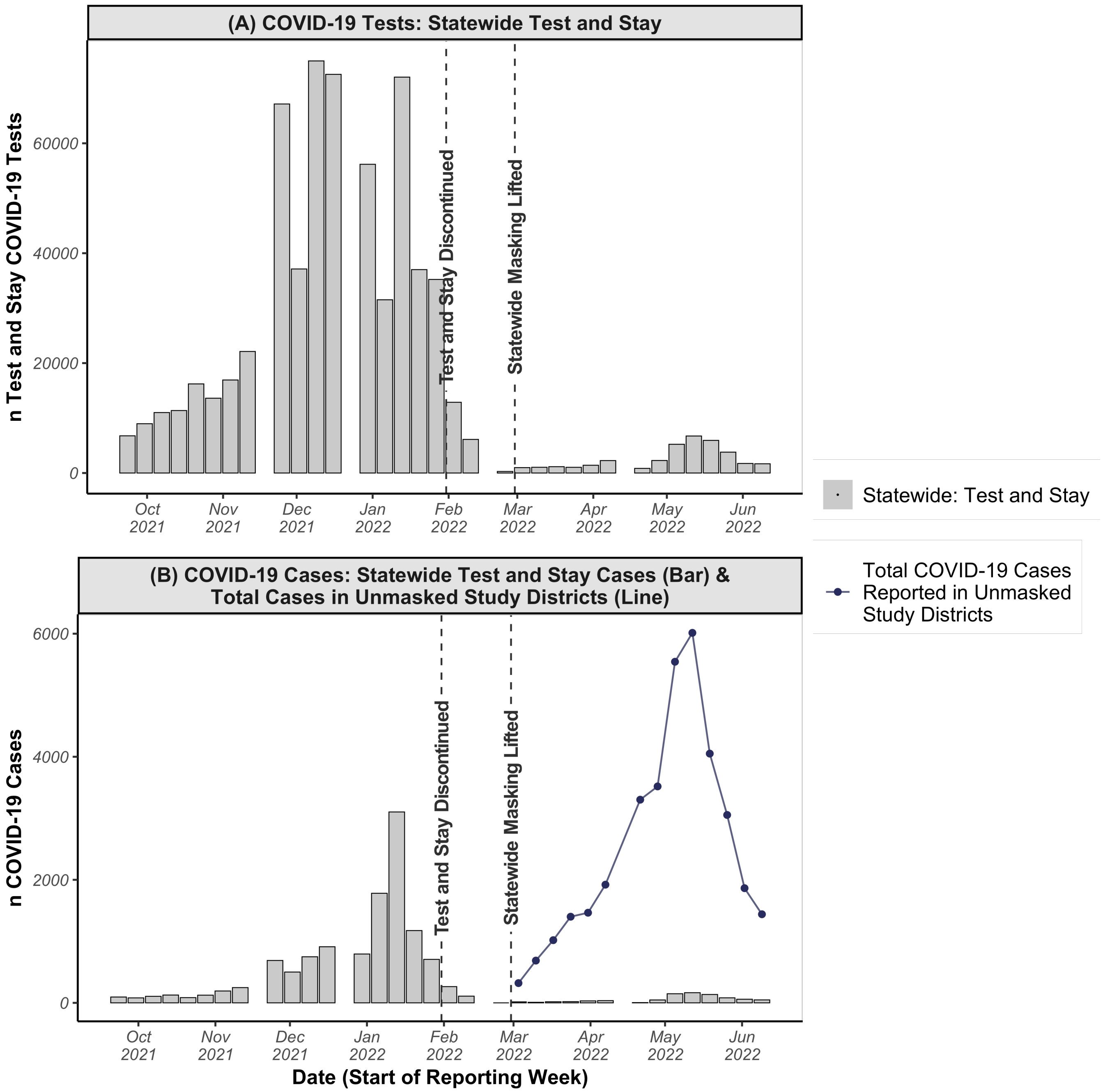
[10] Variables are scaled to enable them to be depicted on the same scale – zero represents the mean value with units in standard deviations

Figure S9. *Weekly number of pooled tests per 100 students by week masking requirements were lifted*



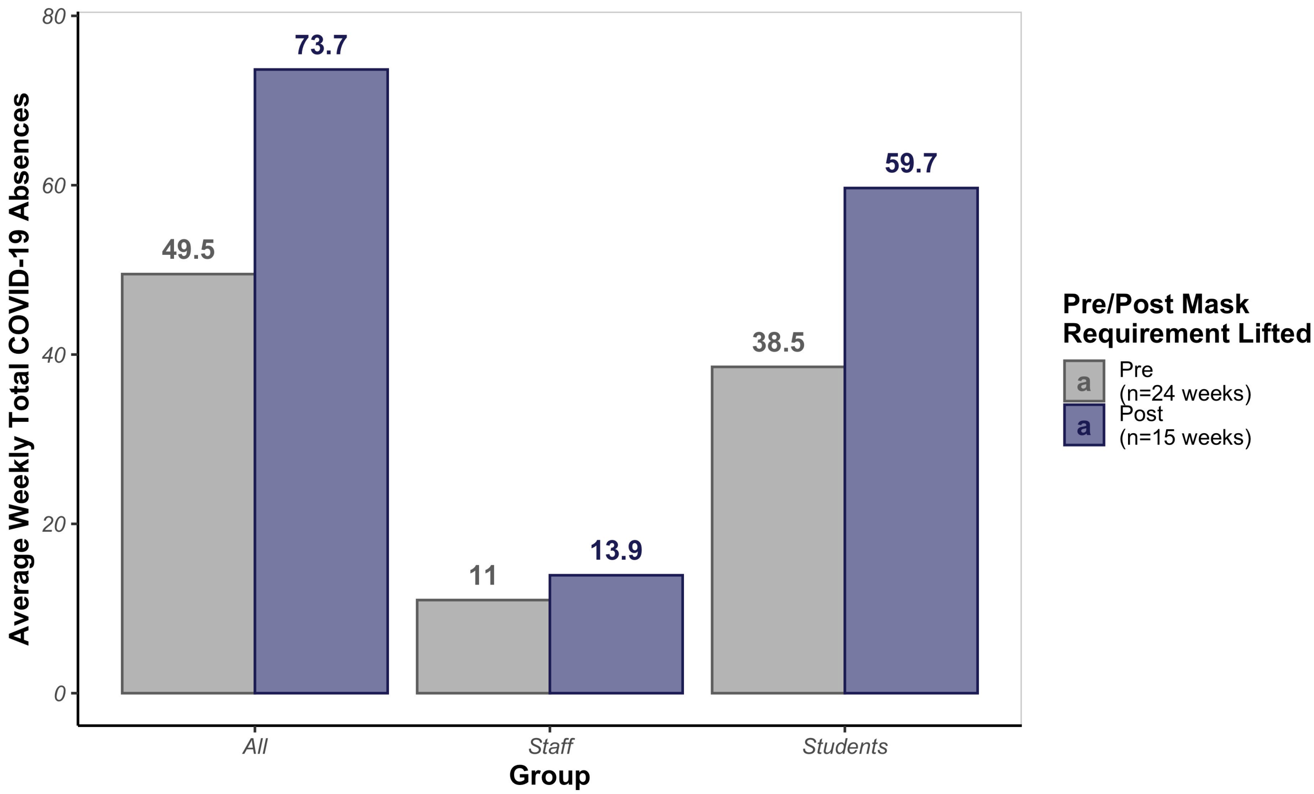
(1) Week of 28-Feb-22: DESE Statewide Mandate lifted (n = 46);
(2) Week of 07-Mar-22 (n = 17);
(3) Week of 14-Mar-22 (n = 7)

Figure S10. *Weekly number of tests (A) and COVID-19 cases identified (B) through DESE's Test and Stay/close contact testing programs statewide*[11] *and total COVID-19 cases identified in our control group study districts before and after DESE recommendation to discontinuation testing of close contacts*



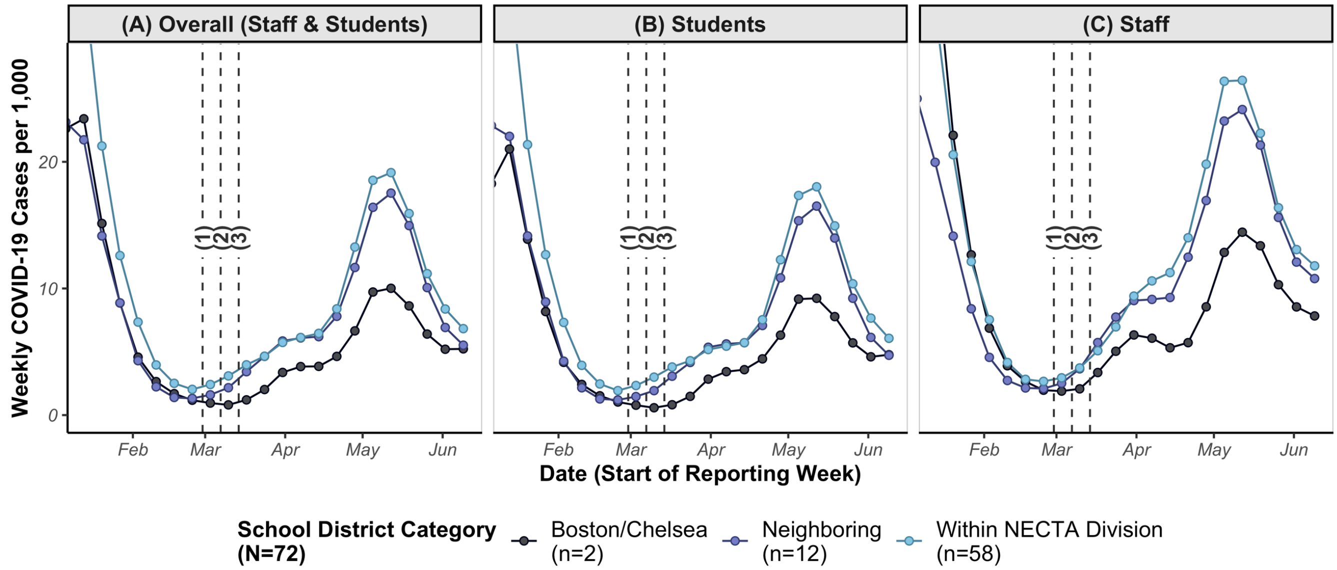
[11] Includes all Massachusetts school districts, both those included and excluded from our study

Figure S11. *Average weekly total absences for students, staff, and overall for the weeks pre- and post-lifting of masking requirements in Lexington Public Schools*[12]



[12] Source: https://sites.google.com/lexingtonma.org/lps-covid19-return-to-school/lps-covid-19-reports-dashboard; Snapshot of LPS for Thursday attendance reported Friday

Figure S12. *Weekly reported rate of COVID-19 cases (A) overall, (B) among students, and (C) among staff in Boston/Chelsea Public Schools, neighboring school districts and non-neighboring districts within the Boston-Cambridge-Newton NECTA division.*[13]



(1) Week of 28-Feb-22: DESE Statewide Mandate lifted (n = 46)
(2) Week of 07-Mar-22 (n = 17)
(3) Week of 14-Mar-22 (n = 7)

[13] Dates on the x-axis were restricted to the period immediately before and after mandates were lifted statewide and in most school districts. Difference-in-differences analysis includes all weeks in the 2021-2022 school year.

SUPPLEMENTARY TABLES

Table S1. *Inclusion and exclusion steps and list of included/excluded school districts in final sample*

<table>
<tr><th>n Remaining</th><th>n Excluded</th><th>Exclusion Step</th></tr>
<tr><td>400</td><td></td><td>Total school districts reporting COVID-19 data to DESE</td></tr>
<tr><td>79</td><td>(-211)</td><td>Exclude charter, vocational, and technical school districts</td></tr>
<tr><td>72</td><td>(-7)*</td><td>Final sample; Exclude unreliable districts; districts were excluded if either:
1. School district had >10 weeks with 0 cases reported among students and staff
OR
2. School district has >5 weeks where either:
a. 0 cases reported among students and staff, but district reported at least 1 positive pool test
OR
b. 0 cases reported among students and staff and ≥10 cases reported on average over the past 4 weeks</td></tr>
<tr><td colspan="3"><u>*Excluded Districts in NECTA (n=7):</u> Arlington, Bedford, Harvard, Scituate, Sherborn, Weston, Watertown

<u>Included districts:</u> Andover, Belmont, Boston, Braintree, Brookline, Burlington, Cambridge, Canton, Carlisle, Chelsea, Cohasset, Concord, Dedham, Dover, Everett, Foxborough, Franklin, Hanover, Hingham, Holbrook, Hull, Lexington, Lincoln, Lynnfield, Malden, Mansfield, Marshfield, Maynard, Medfield, Medford, Medway, Melrose, Millis, Milton, Needham, Newton, Norfolk, North Reading, Norwell, Norwood, Quincy, Randolph, Reading, Revere, Rockland, Saugus, Sharon, Somerville, Stoneham, Stoughton, Sudbury, Wakefield, Walpole, Waltham, Wayland, Wellesley, Westwood, Weymouth, Wilmington, Winchester, Winthrop, Woburn, Wrentham, Acton-Boxborough, Ayer Shirley School District, Concord-Carlisle, Dover-Sherborn, Groton-Dunstable, King Philip, Lincoln-Sudbury, Nashoba, Whitman-Hanson</td></tr>
</table>

Table S2. *Details and dates for Massachusetts Department of Elementary and Secondary Education (DESE) recommended COVID-19 testing programs for the 2021-2022 school year*

Testing Program	Recommendation Period & Description
Symptomatic Testing	**Recommendation Period**: Sep 2021 – Jun 2022 (Full 2021-2022 School Year) **Description**: “(for when individuals present symptoms while at school; individuals should not go to school if experiencing symptoms while at home): Shallow nasal swab samples are collected at school using the BinaxNOW rapid antigen test or another approved diagnostic test.”[14]
Routine Pooled Screening Testing	**Recommendation Period**: Sep 2021 – Jun 2022 (Full 2021-2022 School Year) **Description**: • “Routine COVID Pooled Testing and School-Based Follow-Up Testing: Shallow nasal swab samples are collected at school and put into a single tube (maximum of 10 samples per tube). If a group tests positive, individual Follow-Up testing with a second sample collection occurs at the school with BinaxNOW and/or individual PCR testing, as necessary.” **OR** • “Routine COVID Pooled Testing and Lab-Based Follow-Up Testing: Shallow nasal or saliva samples are collected and kept separate before being grouped at the lab. If a group tests positive, individual Follow-Up testing occurs at the lab, without a second sample collection. Individual test results are reported to the school.”

Test-and-Stay/Close Contact Testing	**Recommendation Period**: Sep 2021 – Jan 2022 (Replaced by weekly At-Home Rapid Testing)
	
	Description: For individuals identified as close contacts, “Shallow nasal swab samples are collected at school using the BinaxNOW rapid antigen test or another approved diagnostic test. Tests are administered daily from the first day of exposure for a duration of five (5) days with testing occurring on school days.” “Close contacts are defined as individuals who have been within 6 feet of a COVID-19 positive, individual while indoors, for at least 15 minutes during a 24-hour period.”
	
	Under the test-and-stay program, “ …the following close contacts are exempt from testing and quarantine response protocols:
	• Asymptomatic, fully vaccinated close contacts: Individuals who are asymptomatic and fully vaccinated are exempt from testing and quarantine response protocols.
	• Masked close contacts in classrooms and on buses: An individual who is exposed to a COVID-19 positive individual in the classroom, so long as the individuals were spaced at least 3 feet apart, or on the bus, while both individuals were masked, is exempt from testing and quarantine response protocols.
	• Close contacts who have had COVID-19 within the past 90 days: An individual who has been previously diagnosed with COVID-19 and then becomes a close contact of someone with COVID-19 is exempt from testing and quarantine response protocols if:
	  ◦ The exposure occurred within 90 days of the onset of their own illness AND
	  ◦ The exposed individual is recovered and remains without COVID-19 symptoms.

Weekly At-Home Rapid Testing	**Recommendation Period**: Jan 2022 – Jun 2022 (Replaced “Test-and-Stay”) **Description**: At-Home Rapid Testing (to be distributed to participating staff and students, regardless of vaccination status): Shallow nasal swab samples are collected at home on a weekly basis using the iHealth rapid at-home test distributed by DESE.”[15]

[14] https://www.doe.mass.edu/covid19/testing/original.html
[15] https://www.doe.mass.edu/covid19/testing/legacy.html

Table S3. *Descriptions of various sensitivity analyses for difference-in-differences analyses. The descriptions here correspond to the results presented in* ***Figure S5****.*

Sensitivity Analysis Category	Type & Details
(A) <u>Data Cleaning Steps</u> *Data cleaning steps address two main challenges – 2-week reporting weeks and missing data reported as zeros.* ● *For n=4 reporting weeks (corresponding to the weeks after Thanksgiving recess, winter recess, February recess, and spring recess), DESE asked that school districts report all cases for the prior two weeks (i.e., reporting period was 2 calendar weeks rather than 1 calendar week).* ● *DESE data does not distinguish between zero cases and non-reporting. Upon examining the data, we noted that often school districts would account for a zero report in the prior week by reporting two weeks' worth of cases in the following week (i.e., school districts were over-correcting for not reporting in the following week).*	**<u>Raw Data:</u>** data as reported by DESE with no changes ● *<u>2-week reporting weeks:</u>* considered data missing for first week, rate calculated over 14 person-days for 2nd week ● *<u>Zero reporting weeks:</u>* Considered true zeros/not corrected
	<u>Correct for 2-week reporting periods:</u> raw data from DESE, but corrects for 2-week reporting weeks ● *<u>2-week reporting weeks:</u>* assumes equal rate across both weeks ● *<u>Zero reporting weeks:</u>* Considered true zeros/not corrected
	<u>Remove non-reporting zeros:</u> Corrects for 2-week reporting periods (as above), and excludes some zeros as non-reporting weeks ● *<u>2-week reporting weeks:</u>* assumes equal rate across both weeks ● *<u>Zero reporting weeks:</u>* Exclude reporting weeks with 0 reported cases across students and staff if: ○ 0 Cases reported, but at least 1 positive pool testing OR ○ 0 cases reported and >10 cases reported on average over the past 4 weeks
	<u>Correct non-reporting zeros:</u> Corrects for 2-week reporting periods (as above), considers some zeros as non-reporting, and assumes over correction in the following week ● *<u>2-week reporting weeks:</u>* assumes equal rate across both weeks ● *<u>Zero reporting weeks:</u>* Exclude reporting weeks with 0 reported cases across students and staff with same rules as above; additionally assume that schools correct for this by reporting prior cases in the following week (i.e., rate in week following zero reporting week is calculated for 2 weeks of person-days)

Sensitivity Analysis Category	Type & Details
(B) <u>Population Weighted</u> *(main analysis)* *Traditional difference-in-differences models treat all observations (i.e., school districts) as having equal weight. Weighting accounts for large variation in school districts size and measures the population-level impact of lifting of masking requirements.*	**<u>Population weighted:</u>** Weekly school district observations are weighted by the school district population size in difference-in-difference models
<u>(C) Smoothed</u> *For similar reasons to (A) above, using rolling averages smooths over the data partially correcting for reporting zeroes instead of missing values.*	**<u>3-week rolling average:</u>** We used a 3-week centered rolling average to minimize noise in the data while preserving the timing of changes in trends (i.e., leading or lagging averages would have shifted the timing of changes in case rates)
<u>(D) Modified Control Group</u> *Our main analysis considers all public-school districts in the Boston-Newton-Cambridge NECTA. In these analyses, we considered different definitions for control counties as school districts more proximate to Boston/Chelsea may serve as better control groups than those further away*	**<u>Primary and secondary neighbors:</u>** Excludes school districts within NECTA that do not share borders Boston/Chelsea or Boston/Chelsea's neighboring school districts
	<u>Primary neighbors only:</u> Excludes school districts within NECTA that do not border Boston/Chelsea
(E) <u>Adjusted for city/town characteristics</u> *Our main analysis did not adjust for any time-dependent covariates; however, we considered these covariates in sensitivity analyses in difference-in-differences models. Models use inverse probability weighting (IPW) for adjustment and build on our main analysis in (B) above.*	**<u>Community COVID-19 Testing Rate:</u>** Adjusted for city or town's total number of reported tests over the past two weeks divided by the city/town population. Data from MA DPH (16-Jun-2022).
	<u>Community COVID-19 Case Rate (Reported):</u> Adjusted for city or town's total number of reported cases over the past two weeks divided by the city/town population. Data from MA DPH (16- Jun-2022).
	<u>Community COVID-19 Case Rate (Corrected):</u> City/towns' COVID-19 case rate adjusted for underreporting using the percent test positivity using methodology from: *Chiu WA, Ndeffo-Mbah ML. Using test positivity and reported case rates to estimate state-level COVID-19 prevalence and seroprevalence in the United States. PLoS Comput Biol. 2021;17(9): e1009374.*[12] *Data from MA DPH (16-Jun-2022).*

Sensitivity Analysis Category	Type & Details
	Community COVID-19 % Test Positivity: Adjusted for city or town's total number of positive tests over the past two weeks divided by the total number of tests over the past two weeks. Data from MA DPH (16-Jun-2022).
	School District Size: *Adjusted for school district population size (i.e., number of students/staff); data from DESE*
	City/Town Population: *Adjusted for city or town population size in which the school district was located. Data from MA DPH (16-Jun-2022).*
	Cumulative School District Case Rate (past 12 weeks): Adjusted for the cumulative number of cases observed in the school district over the past 12 weeks, rolling (i.e., potential measure of prior immunity)
	Cumulative School District Case Rate (start of school year): Adjusted for the cumulative number of cases observed in the school district since the start of the school year (i.e., potential measure of prior immunity)
	Community Vaccination Rate, All Ages: Adjusted for city or town's percentage of all residents fully vaccinated (DPH defined measure for completing primary series) for COVID-19. Data from MA DPH Weekly Municipality Vaccination Reports.
	Community Vaccination Rate, Age 5-11 years: Adjusted for city or town's percentage of residents aged 5-11 years fully vaccinated (DPH defined measure for completing primary series) for COVID-19. Data from MA DPH Weekly Municipality Vaccination Reports.
	Community Vaccination Rate, Age 12-15 years: Adjusted for city or town's percentage of residents aged 12-15 years fully vaccinated (DPH defined measure for completing primary series) for COVID-19. Data from MA DPH Weekly Municipality Vaccination Reports.

Table S4. *Assessment of violations to the parallel trends assumption in the pre-period from event study regression and formal statistical tests*

	Event-Study Results		Cramér-von Mises Statistical Tests		
Group	*Mean (SD) Pre-period ATT[1]*	*n Pre-period ATT[1] Significant*	*Test Statistic*	*Critical Value*	*P-value*
Overall (Students & Staff)	0.01 (1.7)	2 of 26 (8%)	34088.26	52262.61	0.198
Students	0.02 (1.3)	2 of 26 (8%)	37423.8	65487.2	0.241
Staff	0.01 (1.94)	2 of 26 (8%)	60109.34	67844.34	0.101

[1] Weekly Rate Difference in COVID-19 cases per 1,000

Table S5. *Calculation for number of missed school days following 5-day isolation for COVID-19 cases attributable to lifting school masking requirements (See Supplementary Appendix methods)*

	Attributable COVID-19 Cases[16]			**n Missed Days**[17]		
	Estimate	*95% LCL*	*95% UCL*	*Estimate*	*95% LCL*	*95% UCL*
Overall (Staff & Students)	11,901	8,651	15,151	37,242	27,072	47,412
Staff	2,882	2,092	3,673	9,019	*6,547*[18]	11,494
Students	9,168	5,594	12,743	28,690	*17,505*	39,877

[16] Estimates from difference-in-differences models presented in main text **Table 1**

[17] Calculated as: n Attributable Cases x 5 days isolation per case x 0.6259; where 0.6259 represents the proportion of calendar days in school (181.5 school days out of 290 calendar days)

[18] The highlighted and underlined numbers are presented in main text discussion section

Table S6. *Sources and dates that Test-and-Stay/close contact testing programs were discontinued for 10 largest schools that lifted masking requirements (i.e., control group)*

School District	Date Ended	Sources
Newton	Jan 24	https://drive.google.com/file/d/1gWD8Ch0KfP6ax6UjbRQDxZUnsKJH_uhj/view https://www.newton.k12.ma.us/covidtesting
Quincy	Jan 26	https://www.quincypublicschools.com/kindergarten_registration_for_2020-21/1_26_2022_qps_newsletter
Brookline	Jan 27	https://www.brookline.k12.ma.us/site/default.aspx?PageType=3&DomainID=4&ModuleInstanceID=651&ViewID=6446EE88-D30C-497E-9316-3F8874B3E108&RenderLoc=0&FlexDataID=9950&PageID=1
Cambridge	Feb 28	https://www.cpsd.us/district_news/covid_reminders_24feb2022
Revere	Jan 18	https://4.files.edl.io/058a/01/19/22/140845-1eec9280-b00e-4aa7-b3ed-6ad0f2eddded.pdf https://www.reverek12.org/apps/pages/index.jsp?uREC_ID=2191798&type=d&pREC_ID=2188912
Lexington	Feb 2	https://docs.google.com/document/d/11SL6n1lOevLBCFf9h9D64sqfju1EHwuLAAEyXYsc9x4/edit#bookmark=id.bhnn9719hj9s https://docs.google.com/document/d/1_3Oyt3yzwU6-LqPxbHIHD6qSzALGL7ngF81z2TAFzhE/edit
Everett	Jan 31	https://www.everettpublicschools.org/apps/news/article/1570789#:~:text=In%20accordance%20with%20DESE%20and,tracing%20for%20in%2Dschool%20exposures
Malden	Feb 7	https://maldenps.org/wp-content/uploads/2022/02/Superintendents-Report-SC-020722.pdf
Waltham	Jan 31	https://www.smore.com/e96hw-district-monthly-update
Weymouth	Feb 23	https://www.weymouthschools.org/district/district-information/news/covid-testing-information

Table S7. *Percentage of cases occurring pre-Omicron, during the Omicron BA.1 wave, and after masking requirements were lifted*

			Pre-Omicron (Weeks 1-10, n= 10 weeks)		**Omicron BA.1 (Weeks 11-25, n= 10 weeks)**		**Post-Lifting/ Omicron BA.2+ (Weeks 26-40, n=15 weeks)**	
	Week Masking Requirement Lifted	*n Districts*	*n Cases*	*Row %*	*n Cases*	*Row %*	*n Cases*	*Row %*
All	**ALL**	**72**	**5173**	**4.8%**	**62702**	**58.1%**	**40104**	**37.1%**
	Did not lift	2	686	5.1%	8343	62.0%	4437	32.9%
	03-Mar-22	46	2750	5.0%	33941	62.0%	18089	33.0%
	10-Mar-22	17	1212	4.1%	15142	51.0%	13356	45.0%
	17-Mar-22	7	525	5.2%	5276	52.6%	4222	42.1%
Students	**ALL**	**72**	**4405**	**5.0%**	**52146**	**58.9%**	**31948**	**36.1%**
	Did not lift	2	525	5.4%	5824	59.7%	3399	34.9%
	03-Mar-22	46	2388	5.2%	29045	63.7%	14148	31.0%
	10-Mar-22	17	1053	4.2%	12915	51.4%	11142	44.4%
	17-Mar-22	7	439	5.4%	4362	54.1%	3259	40.4%
Staff	**ALL**	**72**	**768**	**3.9%**	**10556**	**54.2%**	**8156**	**41.9%**
	Did not lift	2	161	4.3%	2519	67.8%	1038	27.9%
	03-Mar-22	46	362	3.9%	4896	53.2%	3941	42.8%
	10-Mar-22	17	159	3.5%	2227	48.4%	2214	48.1%
	17-Mar-22	7	86	4.4%	914	46.6%	963	49.1%

Table S8. *Race/ethnicity of students and race/ethnicity and sex of staff for included and excluded school districts*

	Included school districts by week masking requirement was lifted and Excluded School Districts (Outside NECTA)					
Characteristic	03-Mar-22 N = 46	10-Mar-22, N = 17	17-Mar-22, N = 7	Did not lift, N = 2	Not Included, N = 217	Overall, N = 289
N Schools	46 (100%)	17 (100%)	7 (100%)	2 (100%)	217 (100%)	289 (100%)
STUDENTS						
Total Students	**139,752 (100%)**	**69,074 (100%)**	**33,015 (100%)**	**52,243 (100%)**	**534,270 (100%)**	**828,354 (100%)**
Students' Race/Ethnicity						
Native American/ Alaskan Native	285 (0.2%)	107 (0.2%)	66 (0.2%)	140 (0.3%)	1,184 (0.2%)	1,782 (0.2%)
Asian	14,616 (10%)	12,162 (18%)	3,633 (11%)	4,154 (8.0%)	27,921 (5.2%)	62,486 (7.5%)
Black	7,719 (5.5%)	4,132 (6.0%)	4,496 (14%)	13,635 (26%)	38,434 (7.2%)	68,416 (8.3%)
Hispanic/Latinx	16,764 (12%)	6,086 (8.8%)	9,061 (27%)	25,190 (48%)	129,326 (24%)	186,427 (23%)
Multiple	5,417 (3.9%)	4,364 (6.3%)	2,253 (6.8%)	1,669 (3.2%)	22,338 (4.2%)	36,041 (4.4%)
Native Hawaiian/ Pacific Islander	181 (0.1%)	47 (<0.1%)	30 (<0.1%)	74 (0.1%)	372 (<0.1%)	704 (<0.1%)
White	94,731 (68%)	42,165 (61%)	13,476 (41%)	7,381 (14%)	314,491 (59%)	472,244 (57%)
STAFF						
Total Staff	**20,782 (100%)**	**10,660 (100%)**	**5,764 (100%)**	**9,324 (100%)**	**82,280 (100%)**	**128,810 (100%)**
Staff Race/Ethnicity						
Native American/ Alaskan Native	19 (<0.1%)	6 (<0.1%)	4 (<0.1%)	17 (0.2%)	83 (0.1%)	130 (0.1%)
Asian	320 (1.5%)	362 (3.4%)	224 (3.9%)	487 (5.2%)	871 (1.1%)	2,263 (1.8%)
Black	210 (1.0%)	327 (3.1%)	465 (8.1%)	2,466 (26%)	2,180 (2.6%)	5,648 (4.4%)
Hispanic/Latinx	343 (1.6%)	300 (2.8%)	348 (6.0%)	1,369 (15%)	4,266 (5.2%)	6,625 (5.1%)
Multiple	97 (0.5%)	106 (1.0%)	63 (1.1%)	18 (0.2%)	396 (0.5%)	680 (0.5%)
Native Hawaiian/ Pacific Islander	16 (<0.1%)	6 (<0.1%)	6 (0.1%)	13 (0.1%)	47 (<0.1%)	89 (<0.1%)
White	19,778 (95%)	9,553 (90%)	4,654 (81%)	4,954 (53%)	74,439 (90%)	113,377 (88%)
Staff Sex						
Female	16,992 (82%)	8,562 (80%)	4,559 (79%)	7,065 (76%)	66,661 (81%)	103,838 (81%)
Male	3,786 (18%)	2,098 (20%)	1,202 (21%)	2,252 (24%)	15,606 (19%)	24,943 (19%)

REFERENCES

1. Callaway B, Sant'Anna PHC. Difference-in-Differences with multiple time periods. J Econom 2021;225(2):200–30.

2. Callaway B, Sant'Anna PHC. did: Difference in Differences [Internet]. 2021;Available from: https://bcallaway11.github.io/did/

3. Marcus M, Sant'Anna PHC. The Role of Parallel Trends in Event Study Settings: An Application to Environmental Economics. Journal of the Association of Environmental and Resource Economists 2021;8(2):235–75.

4. Callaway B, Sant'Anna PHC. Difference-in-Differences with Multiple Time Periods and an Application on the Minimum Wage and Employment. Working Paper [Internet] 2018;Available from: https://arxiv.org/abs/1803.09015v2

5. Caniglia EC, Murray EJ. Difference-in-Difference in the Time of Cholera: a Gentle Introduction for Epidemiologists. Curr Epidemiol Rep 2020;7(4):203–11.

6. Wing C, Simon K, Bello-Gomez RA. Designing Difference in Difference Studies: Best Practices for Public Health Policy Research. Annu Rev Public Health 2018;39:453–69.

7. Wiens KE, Smith CP, Badillo-Goicoechea E, et al. In-person schooling and associated COVID-19 risk in the United States over spring semester 2021. Sci Adv 2022;8(16):eabm9128.

8. Lessler J, Grabowski MK, Grantz KH, et al. Household COVID-19 risk and in-person schooling. Science 2021;372(6546):1092–7.

9. Johnson KE, Lachmann M, Stoddard M, et al. Detecting in-school transmission of SARS-CoV-2 from case ratios and documented clusters. medRxiv [Internet] 2021;Available from: http://dx.doi.org/10.1101/2021.04.26.21256136

10. Roth J, Sant'Anna PHC, Bilinski A, Poe J. What's Trending in Difference-in-Differences? A Synthesis of the Recent Econometrics Literature [Internet]. arXiv [econ.EM]. 2022;Available from: http://arxiv.org/abs/2201.01194

11. DESE/DPH Protocols for Responding to COVID-19 Scenarios - SY 2021-22 [Internet]. 2022 [cited 2022 Jun 22];Available from: https://www.doe.mass.edu/covid19/on-desktop/protocols/protocols.pdf

12. Chiu WA, Ndeffo-Mbah ML. Using test positivity and reported case rates to estimate state-level COVID-19 prevalence and seroprevalence in the United States. PLoS Comput Biol 2021;17(9):e1009374.