Statistical Analysis Plan

Statistical analysis plan for: Cowger TL, Murray EJ, Clarke J, et al. Lifting universal masking in schools — Covid-19 incidence among students and staff. N Engl J Med 2022;387:1935-46. DOI: 10.1056/NEJMoa2211029

STATISTICAL ANALYSIS PLAN

Cowger TL, Murray EJ, Clarke J, et al. Impact of Lifting Universal School Masking Requirements on Incidence of COVID-19 among Staff and Students in Greater-Boston Area School Districts: A Difference-in-Differences Analysis. August 2022.

This supplement contains the following items:

Statistical analysis plan (only one version) | *Date Compiled:* 26-Aug-2022

TABLE OF CONTENTS

INTRODUCTION AND AIMS

Alongside other preventive measures, masking represents an important piece of a layered mitigation strategy to reduce SARS-CoV-2 spread in schools and secondary transmission to household members and the surrounding community.[1–4] Massachusetts (MA) was one of 18 states plus DC with statewide school masking requirements during the 2021-2022 school year.[50] The Massachusetts Department of Elementary and Secondary Education (DESE) lifted statewide school masking requirements on February 28, 2022. Subsequently, many MA school districts lifted masking requirements immediately, several retained masking requirements for several weeks, and only two districts – Boston and neighboring Chelsea Public Schools – maintained masking requirements through June 2022.

The staggered lifting of school masking requirements across greater-Boston school districts presented a unique opportunity to evaluate the potential impact of universal school masking policies during a period of highly transmissible SARS-CoV-2 variants in a dynamic COVID-19 policy environment.

Specifically, this study aims to:

(1) Compare weekly incidence of COVID-19 in school districts that lifted masking requirements to districts where masking requirements had not yet been lifted;
(2) Estimate the difference in COVID-19 incidence among students and staff attributable to lifting masking protections; and
(3) Compare school district characteristics for districts that chose to sustain masking policies for longer periods of time to those that lifted masking requirements earlier.

METHODS

Data Sources & Definitions

We will use publicly-available, school district-level data on weekly COVID-19 cases, enrollment, and staffing for the 2021-2022 school year from DESE.[51,52] Throughout the study period, DESE required standardized weekly reporting of all students and staff who tested positive for COVID-19, regardless of symptoms, testing type or program (e.g., symptomatic testing vs. pooled PCR testing), or testing location (community vs. in-school testing) (**Table 1**). DESE also strongly recommended and provided full funding for districts to opt-in to standardized state-sponsored school testing programs and approximately 95% (n=2,311) MA schools participated in at least one such testing program.

School district websites and/or local news sources provided dates of school district masking policies included in this study. In addition, we obtained data on city/town-level COVID-19 indicators and age-specific vaccination rates from the Massachusetts Department of Public Health (DPH) for covariate adjustment in sensitivity analyses. Descriptive analyses used DESE data[52] on school district characteristics, including student/staff sociodemographics, enrollment of DESE-defined selected populations (e.g., low-income students, English learners, students with disabilities) and the 2016 School Survey (most recent data) from the Massachusetts School Building Authority (MSBA) for school building conditions and learning environment.[55]

Case Ascertainment & 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 in the Inclusion/Exclusion criteria section and in **Table 2** 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 1**). 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 (See **Table 1**). 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 Methods section '*Robustness of results to differences in testing programs*,' below.

Exposure and Outcomes

Our primary exposure was maintenance or lifting of masking requirements in each reporting week. All schools had masking requirements in place at the start of our study through February 28, 2022 (i.e., before statewide requirement was lifted). A school district was considered to have lifted their masking requirement if their policy was rescinded prior to the first day of the reporting week. Our primary outcome was weekly rates of COVID-19 per 1,000 staff and students considered together and separately.

Inclusion/exclusion criteria

As summarized in **Table 2** 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 **Figure 1**). 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 (See **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).

Statistical 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.[9–12] Briefly, DiD methods allow estimating the causal effect of policy changes enacted at the group-level by comparing changes in the outcome over time between intervention and control groups, under an assumption of parallel trends (i.e., in the absence of intervention, outcomes in the intervention and control groups would have remained parallel over time).[11,13,14] Unlike some observational methods, 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. In this analysis, we estimated the weekly and cumulative impact over 15 weeks of removing mask requirements on COVID-19 cases in schools that removed masking requirements (i.e., average treatment effect among the treated).

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,[11,15] 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,[11,15] 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,[11] however briefly, `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:

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

- 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 same `did` package, we employed two different aggregation schemes to summarize these group-time average treatment effects – event study/dynamic effects aggregation and calendar time aggregation (see Callaway and Sant'Anna 2021).[11] Event study aggregation uses a weighted-average of $ATT(g,t)$ estimates grouped by length of time since lifting of masking requirements. In event study aggregation, 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 weeks post-lifting of the statewide requirement. Consistent with Callaway and Sant'Anna 2021,[11] we first calculated a weekly measure as a weighted average across all groups that have lifted their masking requirements by time, t, as

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

---

[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).

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:

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

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.[11]

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[13,14,16], following the methods outlined in Callaway & Sant'Anna, 2018.[1,2,18]

Limited Treatment Anticipation Assumption

In addition to parallel trends assumption, identification of causal effects requires an assumption of limited treatment anticipation.[11] 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., 0 weeks anticipation) to anticipation periods of 1,2, and 3 weeks prior to lifting if masking requirements.

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.[19] 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 3**.

Sensitivity Analyses

We conducted several sensitivity analyses to assess whether our results were robust to various data cleaning steps and model specifications. These included a formal test for parallel pre-trends in DiD analyses, adjustment for time-varying covariates including measures of community-level COVID-19 burden, vaccination rates, and rates of prior infection, and assessing the potential impact of differences in testing definitions/programs between districts 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 3**, below. Our final analysis corrected for non-reporting as zeros, considered all school districts within the NECTA division as comparators, and was weighted by school population size to capture the population impact of masking policies.

Descriptive Analysis

Finally, to help understand factors underlying COVID-19 policy decisions in schools and their potential to exacerbate or mitigate COVID-19 inequities, we conducted a descriptive analysis comparing timing of lifting/retaining school masking policies across various school district characteristics, including sociodemographics and physical characteristics of the learning environment.

Calculation of school days missed (presented in discussion)

To calculate the number of missed school days attributable to lifting of mask requirements presented in the discussion, we started with the estimated number of cases attributable to lifting of the masking requirements. 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.

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.

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

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 underdetection/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.

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."[12] Therefore, <u>prior to the lifting of the statewide mandate</u>, 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. 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 <u>will no longer be identified as close contacts</u> by school health professionals." (Page 5, <u>Jan 18 guidance</u>)*[12]

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. We assessed statewide participation in Test and Stay over time to see if at a state-level, Test-and-Stay programs were largely discontinued. To assess whether our results could be due to differences in testing of masked and unmasked close contacts.

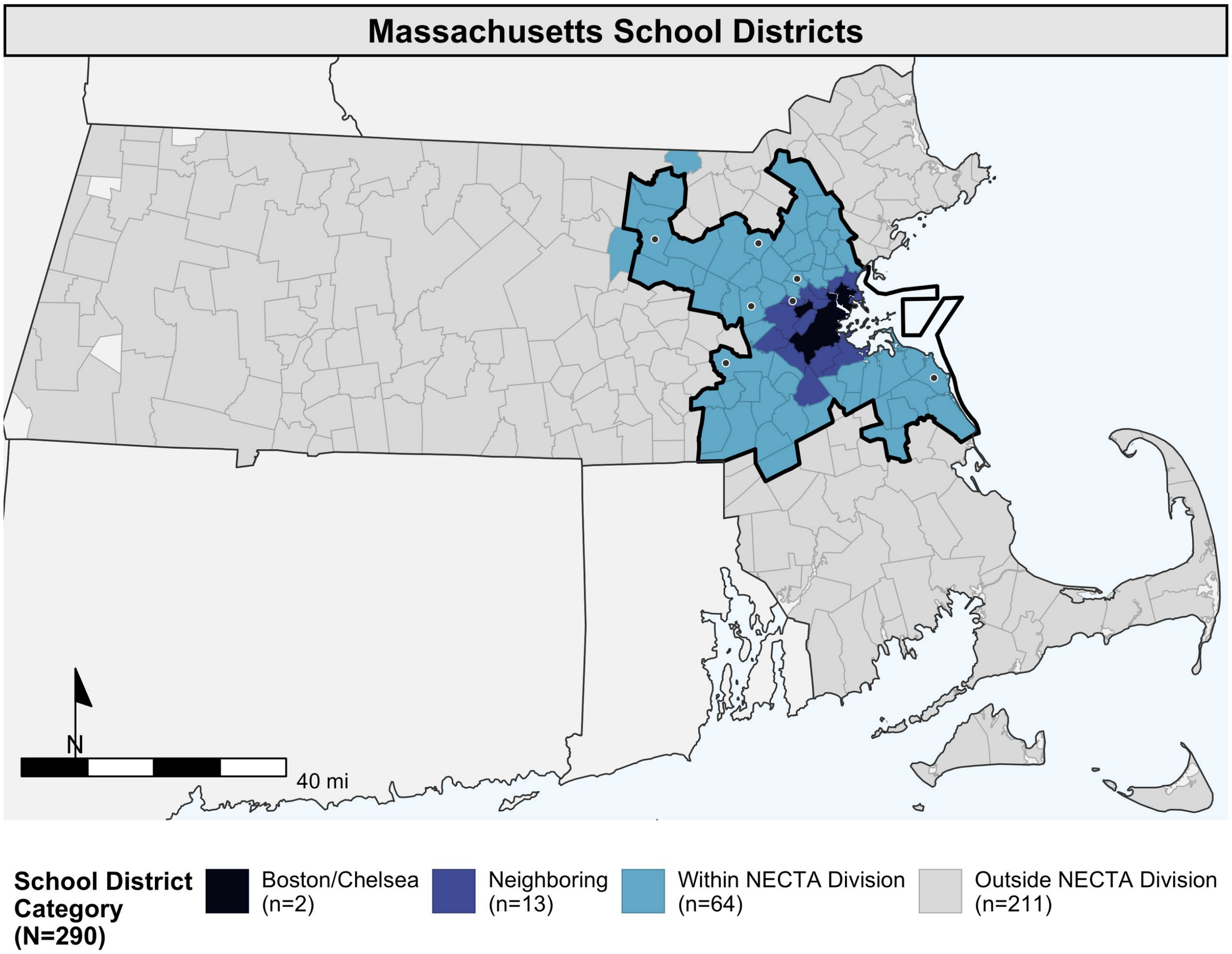
In the most extreme case, if we assume:

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, we compared the total number of cases identified through Test-and-Stay to the number of cases we estimate were attributable to lifting of masking requirements. In addition, we reviewed publicly available data for the 10 largest school districts in our study, to assess whether these districts continued Test-and-Stay.

FIGURES

Figure 1. *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

TABLES

Table 1. *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."[4]
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.

[4] https://www.doe.mass.edu/covid19/testing/original.html

Testing Program	Recommendation Period & Description
	• 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."[5]

[5] https://www.doe.mass.edu/covid19/testing/legacy.html

Table 2. *Inclusion and exclusion steps and list of included/excluded school districts in final sample*

n Remaining	n Excluded	Exclusion Step
400		Total school districts reporting COVID-19 data to DESE
79	(-211)	Exclude charter, vocational, and technical school districts
72	(-7)*	**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
***Excluded Districts in NECTA (n=7):** Arlington, Bedford, Harvard, Scituate, Sherborn, Weston, Watertown **Included districts:** 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		

Table 3. *Descriptions of various sensitivity analyses for difference-in-differences analyses.*

Sensitivity Analysis Category	**Type & Details**
(A) Data Cleaning Steps *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).*	**Raw Data:** data as reported by DESE with no changes ● *2-week reporting weeks:* considered data missing for first week, rate calculated over 14 person-days for 2nd week ● *Zero reporting weeks:* Considered true zeros/not corrected
	Correct for 2-week reporting periods: raw data from DESE, but corrects for 2-week reporting weeks ● *2-week reporting weeks:* assumes equal rate across both weeks ● *Zero reporting weeks:* Considered true zeros/not corrected
	Remove non-reporting zeros: Corrects for 2-week reporting periods (as above), and excludes some zeros as non-reporting weeks ● *2-week reporting weeks:* assumes equal rate across both weeks ● *Zero reporting weeks:* 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
	Correct non-reporting zeros: Corrects for 2-week reporting periods (as above), considers some zeros as non-reporting, and assumes over correction in the following week ● *2-week reporting weeks:* assumes equal rate across both weeks ● *Zero reporting weeks:* 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

Sensitivity Analysis Category	Type & Details
	following zero reporting week is calculated for 2 weeks of person-days)
(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>	**<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).

Sensitivity Analysis Category	Type & Details
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.	**Community COVID-19 Case Rate (Reported):** 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).
	Community COVID-19 Case Rate (Corrected): 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.*[13] *Data from MA DPH (16-Jun-2022).*
	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).
	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)
	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).
	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

Sensitivity Analysis Category	Type & Details
	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.

REFERENCES

1. 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.

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

3. Gettings J, Czarnik M, Morris E, et al. Mask Use and Ventilation Improvements to Reduce COVID-19 Incidence in Elementary Schools - Georgia, November 16-December 11, 2020. MMWR Morb Mortal Wkly Rep 2021;70(21):779–84.

4. Falk A, Benda A, Falk P, Steffen S, Wallace Z, Høeg TB. COVID-19 Cases and Transmission in 17 K-12 Schools - Wood County, Wisconsin, August 31-November 29, 2020. MMWR Morb Mortal Wkly Rep 2021;70(4):136–40.

5. Decker S. Which States Ban Mask Mandates in Schools, and Which Require Masks? [Internet]. Education Week. 2021 [cited 2022 Jul 5];Available from: https://www.edweek.org/policy-politics/which-states-ban-mask-mandates-in-schools-and-which-require-masks/2021/08

6. Massachusetts Department of Elementary and Secondary Education. Positive COVID-19 Cases in Schools [Internet]. [cited 2022 Jun 22];Available from: https://www.doe.mass.edu/covid19/positive-cases/default.html#weekly-report

7. School and District Profiles - Statewide Reports [Internet]. Massachusetts Department of Elementary and Secondary Education (DESE). Available from: https://profiles.doe.mass.edu/state_report/

8. Massachusetts School Building Authority. 2016 SCHOOL SURVEY REPORT [Internet]. 2016. Available from: https://www.massschoolbuildings.org/sites/default/files/edit-contentfiles/Programs/School_Survey/2016/MSBA%202016%20Survey%20Report_102417-FINAL.pdf

9. Athey S, Imbens GW. Design-based analysis in Difference-In-Differences settings with staggered adoption. J Econom 2022;226(1):62–79.

10. Goodman-Bacon A. Difference-in-differences with variation in treatment timing. J Econom 2021;225(2):254–77.

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

12. 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

13. 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.

14. 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.

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

16. 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.

17. 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

18. 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

19. 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

20. 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.