Manuscript deliverable: School district policy guidance as a tool for mitigating the spread of COVID-19 among schools

# Introduction

The purpose of this analysis is to study the effectiveness of school districts’ COVID-19 prevention strategies in mitigating the spread of COVID-19 among school-aged children and adolescents.

As such, the primary research question is, what is the association between districts’ COVID-19 prevention policies and subsequent COVID-19 caseloads in schools?

# Methods

## Data

This research utilizes district policy guidance and school-level COVID-19 case counts collected as part of the National School COVID-19 Prevention Study (NSCPS), funded by the Centers for Disease Control and Prevention (CDC). The NSCPS is a nationally representative stratified random sample of 1,602 K-12 public schools across 1,286 districts, drawn to better understand schools’ response to the pandemic through the 2021-2022 school year and associated outcomes, including the extent to which prevention strategies were effective in mitigating the spread of COVID-19 among students. As part of the NSCPS, publicly-posted guidance documents from district websites were scraped and analyzed to identify prevention strategies that were recommended or required of schools. In addition, health departments representing all 50 states and the District of Columbia were contacted and invited to provide any available COVID-19 case count data for the 2021-2022 school year that had been collected from schools comprising the sample.

## Measures

**Outcome of interest**: The outcome of interest is the difference between schools’ spring and fall monthly average case counts, characterized as the average of schools’ monthly number of cases per 100 students across January, February, and March 2022 minus the average of October, November, and December 2021 monthly cases per 100 students. This outcome is hereafter referred to as ‘change in case rate.’

**Independent variables:** The key predictors of interest consist of 10 dichotomous indicators of COVID-19 mitigation strategies aligned as stringently with CDC guidance as possible[ADD FURTHER DESCRIPTION OF CDC GUIDANCE AND HOW WE ALIGNED WITH IT]. Mitigation strategies were considered in place if they had been updated on the district website during the fall of 2021. These 10 strategy indicators include:

* *Vaccination offered*: Offered vaccines at district-sponsored events to teachers and staff and/or students.
* *Universal masking requirement*: Teachers, staff, and students required to wear masks consistently and correctly (i.e., covering the mouth and nose) at school.
* *Physical distancing*: Students maintain at least 3 feet of physical distance between each other indoors.
* *Screening testing for students*: Encouraged/recommended/offered screening testing of students on a regular basis.
* *Staying home when sick*: Encouraged or recommended that students stay home when sick or tested positive for COVID-19.
* *Contact tracing*: Encouraged or recommended that schools conduct contact tracing.
* *Quarantining*: Required students to quarantine if identified to be a close contact.
* *Cleaning*: Required schools to clean high touch surfaces at least once a day or between uses.
* *HEPA filters*: Encouraged/recommended/offered use of high-efficiency particulate air (HEPA) filters.
* *HVAC systems*: Encouraged/recommended/offered replacing, upgrading, maintaining, or inspecting HVAC systems.
* *Cumulative strategy index*: Sum of strategies having marginal (p-value < .10) association with change in case rates.

*School-level characteristics:* These measures were derived from the National Center for Education Statistics (NCES) for the 2020-2021 school year, and when possible, missing values were filled with estimates from the 2019-2020 school year. See next section for description of approach to ameliorate potential of issues with respect to multicolinearity.

* *Percent student body eligible for free and reduced lunch*: The percent of the schools’ students who were eligible or free and reduced lunch
* *School locale*: City, Rural, Suburb, Town, Missing
* *Percent of student body Asian, American Indian or Alaska Native, Black/African American, Hispanic/Latino, Native Hawaiian or other Pacific Islander, Not specific, Two or more races, and White (each race/ethnicity represented individually)*

*County-level characteristics*

* *Social Vulnerability Index (SVI)*: Overall summary index indicating the relative vulnerability of U.S. Census tracts across four themes: socioeconomic, household composition & disability, minority status & language, and housing type & transportation. Drawn from the….
* *Change in county COVID-19 case rates*: Difference in average of 7-day rolling average for the 15th of each month case rate per 100,000 people between October - December and January - March, corresponding with the time period used for calculating school case rate changes. Pulled from HHS Protect
* *Region*: Midwest, Northeast, South, West
* *State*: 20 states (should we list them?)

## Analyses

In total, policy documents from 1,184 of 1,286 (92%) of the districts were collected, with 28 of the 51 health departments (55%) reporting on 641 schools (40% of sample). Schools without at least one month of reporting during the spring and fall periods were dropped, resulting in an initial sample of 347 schools (22%) across 338 districts (26% of districts) and 20 states (39%).

All analyses were conducted using R version 4.2.1. The analysis begins by testing for outliers, identified as any observation in which the change in case rate is outside 3.5 standard deviations from the mean. Descriptive statistics for all study variables are reported, as well as t-tests of mean differences between groups defined by the presence of mitigation strategy, and Pearson’s correlation coefficient between changes in case rates and continuous covariates following standardization. Finally, intra-class correlation coefficients (ICC) are calculated to test for clustering of outcomes by region, state, and district. For modeling results, 95% confidence intervals and p-values are displayed, with a significance threshold of and marginal significance indicated .

Following the descriptive analysis, the modeling sequence consists of two stages. First, developing Random Forest (RF) algorithms for identifying the most predictive covariates among the school-level variables (student population composition, free lunch eligible, etc.). RF algorithms can be used to rank variables based on their predictive association with the outcome of interest, and due to limited sample size for the current study as well as collinearity between school-level variables, we elected to utilize a data-driven approach for a priori excluding covariates with the least predictive value. Covariates with positive variable importance for greater than 50% of the 100 iterations were retained for modeling.

The second stage of analysis is comprised of three sets of multilevel models, accounting for nesting of schools within state, nested within region. First, individual multilevel models with one strategy and important covariates as predictors are estimated to assess each strategy’s association with the change in case rates before inclusion of other strategies; secondly, estimation of a full model including all strategies for relative comparison; and third a model to compare schools having multiple strategies in place, characterized by a cumulative index. The cumulative index is calculated as the sum of strategies identified form the first set of models as having a marginal association (p-value < .10) with change in case rates.

# Results

Five schools had changes in case rates greater than 3.5 standard deviations from the mean, and were subsequently removed from the dataset for all analyses, resulting in an initial sample of 342 schools. Table 1 provides summary statistics. Three hundred forty-two schools had case data available for fall and spring, with an overall average of 1.16 (SD = 1.87) more cases per month during the spring of 2022 than fall 2021. Surprisingly, none of the school-level covariates were significantly associated with changes in case rates. By region, the final sample consisted of 55 schools from the Midwest (16%), 94 from the Northeast (27%), 120 from the South (35%), and 73 from the West (21%). Schools were comparatively distributed by locale, including city (24%), rural (27%), suburb (35%), and town (14%). Table one also displays minimum, maximum, mean, and standard deviation of case rates by region and locale. For region, state, and locale, ICCs indicated significant clustering by region (ICC = .08) and state (.22), though not by district (.00). As such, accounting for clustering of schools by region and state was necessary for modeling.

Table 2 reviews summary statistics and t-test results of the 10 prevention strategies. *No policy* provides the change in case rate among schools not having the policy in place, whereas *Has policy* indicates the change in case rate among schools having the policy. *Difference in means* provides the mean difference between groups, along with p-values from t-tests, with a signficance threshold of . Strategies were associated with smaller increases in case rates between semesters, though significant differences were only detected for HVAC systems (mean difference = .84; p-value = .02), with a marginally significant difference for physical distancing (mean difference = .88; p-value = .08).

Table three shows results from the first set of multilevel models run individually for each strategy. School-level covariates include (but are suppressed) percent student body two or more races, Asian, White, Black or African American, percent free and reduced lunch, school level, and county-level indicators include change in COVID-19 case count rate and SVI Overall Rank. None of the strategies were statistically significant, though three demonstrated marginal association, including physical distancing (coefficient = -.38; p-value = .07), staying home when sick (-.33; .08), and HVAC systems (-.38; .06). These three strategies were selected for calculation of the cumulative index reviewed below.

Table four shows multilevel model results from including all strategies as predictors in one model. Overall, none of the strategies were significantly associated with changes in case rates. However, percent of student body two or more races was associated with increased changes in case rates (.24; 95% CI = .04 - .46).

Table five shows results from the multilevel model including the cumulative index as the predictor if interest. Given that three strategies were selected for inclusion in the sum score, the index was treated as ranked ordinal. As such, the baseline schools were those not having any strategies in place. Schools with all three strategies in place had a significantly smaller increase in case rates between fall and spring (-.53; -1.07 - -.05), and percent student body two or more races retained statistical significance as well (.24; .03 - .46). Finally, percent student body on free and reduced lunch was associated with a marginally significant decrease (-.28; -.58 - .05).

# Discussion

Overall, these results suggest that districts providing mitigation strategy guidance to schools can reduce the transmission of COVID-19 among school-aged children and adolescents. These results also suggest that, individually, mitigation measures may not be enough to reduce the spread of COVID-19, and rather that mitigation requires combinations of strategies to be utilized.

Write- how is our study different? We look at district guidance and whether it translates to results on the ground. We also compare strategies to assess whether they can individually reduce transmission. As opposed to youth development interventions where implementation fidelity is less important and tailoring to the local context takes precedence, this may be a case where tailoring is limited to ensuring that the set of strategies are embraced.

Write–how might this be advantageous to the CDC and government for better disseminating guidance?

Does this mean that for reducing transmission, we need to be less worried about tailoring interventions to the local environment, and ensuring more that all elements are used?