### Analysis for Approach 2

#### Methods

The second analytic approach used to understand the association between implementation of COVID-19 prevention strategies and COVID-19 cases utilizes uses data from school district COVID-19 policies scraped from websites and school-level COVID-19 cases counts collected from health departments.

As part of the NSCPS, publicly available school district COVID-19 guidance (e.g., policies, reopening plans, prevention guidance) were retrieved from district websites and analyzed to identify prevention strategies that were required or recommended at the district-level. In addition, health departments representing all 50 states and the District of Columbia were contacted and invited to provide available school-level COVID-19 case count data for the NSCPS sample of schools (N=1,602) for the 2021–2022 school year.

#### Measures

##### Dependent variable

For this analysis, the dependent variable was defined as the difference between schools’ spring and fall monthly average of incident COVID-19 case counts, characterized as the monthly average of school-level number of cases per 100 students from January through March of 2022 minus the monthly average of number of cases per 100 students from October through December of 2021. This outcome is hereafter referred to as the “change in the school-level case rate.”

for school i.

*Independent variables*

Predictors of interest consisted of school district COVID-19 policies, based on scoring of published prevention strategy guidance that were aligned to represent adherence to CDC’s operational guidance for K-12 schools to support safe in-person learning during the COVID-19 pandemic.

School district policies were scored using a combination of human scoring and machine learning (ML) methods. Human raters met to standardize how to categorize and rate COVID-19 prevention guidance using a 41-item scoring rubric assessing the extent to which school district policies required or recommended implementation of 10 key COVID-19 prevention strategies (Table 25). Once raters agreed on the requirements and scoring criteria, they reviewed 427 policy guidance documents containing image files and infographics and recorded their scores. Criteria for scoring were also used to train a keyword matching model ML algorithm that was used to assess and score 757 policy guidance documents in pdf and Microsoft Word format.

*Covariates*

A number of school-level and count-level characteristics of interest were selected as covariates based on XYZ. Please refer to Table 25 for the list of covariates and operational definitions of each.

Table 25. Study measures

| Measure | Definition |
| --- | --- |
| ***Covid-19 Prevention Strategies*** |  |
| 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 | Required that students maintain at least 3 feet of physical distance between each other indoors. |
| Screening testing for students | Offered screening testing of students on a regular basis. |
| Staying home when sick | Encouraged or required 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 use of high-efficiency particulate air (HEPA) filters. |
| HVAC systems | Encouraged replacing, upgrading, maintaining, or inspecting HVAC systems. |
| Cumulative strategy index | Sum of strategies having marginal (p < .10) association with change in case rates. |
| ***School-level characteristics1*** |  |
| Percent student body eligible for free and reduced lunch | The percent of the schools' students who were eligible for free and reduced lunch |
| School locale | City, rural, suburb, town |
| Study enrollment composition | 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 | Overall summary index indicating the relative vulnerability of U.S. Census tracts across four themes: socioeconomic, household composition and disability, minority status and language, and housing type and transportation (Figure 35). Drawn from the American Community Survey (ACS) of the U.S. Census Bureau |
| 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 |

#### 1 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 analyses for description of approach to ameliorate potential of issues with respect to multicollinearity.

In total, the ICF team collected policy guidance documents from 1,186 of 1,271 (93%) school districts from the NSCPS total sample. School-level COVID-19 case data were collected from 28 of 51 health departments (55%) which accounted for 641 schools (40% of 1602 in the total sample). A complete description of case data findings are presented in Appendix XX. Schools without at least one month of case reporting during fall 2021 and spring 2022 were excluded from the analytic sample, resulting in a final analytic sample of 502 schools (31% of total sample of NSCPS schools) across 388 school districts (31% of total sample of NSCPS school districts and 20 states (39% of states and DC).

All analyses were conducted using R version 4.2.1 and included testing for outliers, identified as any observation in which the change in case rate is outside 3.5 standard deviations from the mean (R Core Team, 2022). Descriptive statistics for all study variables were computed (ns, mean, min, max, standard deviation), as well as t-tests of mean differences between groups defined by the presence of prevention strategy guidance (inclusion/adherence to federal guidance) and Pearson’s correlation coefficient between changes in school-level COVID-19 case rates and continuous covariates following standardization. Intra-class correlation coefficients (ICC) were calculated to estimate 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 by p-value .

Following the descriptive analysis, the modeling sequence consisted of two stages. First, a Random Forest (RF) algorithm was developed for identifying covariates most predictive of the change in the school-level COVID-19 case rate among the school-level variables (e.g., student population composition, free lunch eligible, etc.), and completing this process for over 100 iterations (Strobl et al. 2008). RF algorithms can be used to rank variables based on their predictive association with the outcome of interest. Due to the limited sample size for this analysis as well as expected collinear relationships between school-level covariates, the ICF team elected to utilize a data-driven approach for a priori excluding covariates with the least predictive value (Breiman, 2001). For the current analysis, covariates with positive variable importance for greater than 50% of the 100 iterations were retained for subsequent modeling.

The second stage of analysis involved a sequential approach to building multilevel models, using the lme4 package to build models, all accounting for random effects of state and region, and with the default variance-covariance structure of unstructured (Finch, Bolin, & Kelley 2019). The first set of models examined adjusted associations between each school-district COVID-19 policy and the change in the school-level case rate, adjusting for cvoariates. The second set of models examined adjusted associations between each school-district COVID-19 policy, after adjustment for all school-district COVID-19 policies and covariates. Next, a set of models was built to compare schools having multiple strategies in place, characterized by cumulative indices. The cumulative indices were calculated as the sum of strategies identified from the first set of models (estimating association between one strategy and change in school-level case rate) that had p-values less than 0.1 and 0.3. For example, the first of these cumulative indices was comprised of individual strategies whose association with change in school-level case rates yielded p-values less than 0.1.

Results for Approach 2

Ten schools had changes in case rates greater than 3.5 standard deviations from the mean and were subsequently removed from the data for these analyses, resulting in a sample of 502 schools. Table 26 provides summary statistics for this sample. For the continuous covariates (school composition) numbers reflect percentage. For example, the minimum percentage of student body that was American Indian/Alaska Native was 0%, and max 98.7%. The mean percent of student body being AI/AN was 1.47%, with a standard deviation of 8.18%. Five hundred and two schools had case data available for fall 2021 and spring 2022, with an overall average of 1.33 (SD = 2.07) more cases per 100 students per month during spring 2022 than fall 2021. Five of the school-level covariates and county-level predictors were significantly associated with changes in case rates, including county change in case rate (correlation = .19, p-value = 0.00), percent Asian (correlation = .13, p-value = .00), percent Hispanic or Latino (correlation = .15, p-value = .00), percent White (correlation = -.13, p-value = .01), and SVI (correlation = .11, p-value = .01. ICCs indicated significant clustering by region (ICC = 0.07) and state (ICC = 0.22), though not by district (ICC = 0.01). Therefore, accounting for clustering of schools by region and state was necessary for the modeling stage. Overall, the final sample consisted of 69 schools from the Midwest (14%), 130 from the Northeast (26%), 199 from the South (40%), and 104 from the West (21%). Schools were comparatively distributed by locale, including city (33%), rural (22%), suburb (34%), and town (11%). For region and locale, table 2 displays minimum, maximum, mean, and standard deviation of case rates. For standarization of continuous covariates, z-scores were calculated using mean and standard deviation.

Table 26. Summary statistics of changes in case rates and standardized covariates

| Construct | n (min, max) | Mean (SD) | Correlation  (p-value) |
| --- | --- | --- | --- |
| Change in school COVID-19 case rate | 502 (-6.25, 8.75) | 1.33 (2.07) |  |
| Change in county COVID-19 case rate | 502 (−27.72, 14747.43) | 1536.05 (3303.78) | 0.186 (0.000) |
| Percent American Indian/Alaska Native | 498 (0, 98.7) | 1.47 (8.18) | −0.041 (0.356) |
| Percent Asian | 498 (0, 75.8) | 4.77 (9.35) | 0.134 (0.003) |
| Percent Black or African American | 498 (0, 100) | 15.04 (22.46) | -0.038 (0.402) |
| Percent Hispanic or Latino | 498 (0, 100) | 30.02 (28.78) | 0.146 (0.001) |
| Percent Native Hawaiian or other Pacific Islander | 498 (0, 9.7) | 0.26 (0.70) | -0.012 (0.793) |
| Percent no race specified | 498 (0, 2.6) | 0.02 (0.14) | −0.046 (0.309) |
| Percent two or more races | 498 (0, 23.8) | 3.66 (3.09) | −0.045 (0.32) |
| Percent White | 498 (0, 100) | 44.78 (32.8) | −0.126 (0.005) |
| Percent free or reduced price meals | 490 (0, 100) | 54.29 (28.24) | 0.066 (0.144) |
| SVI Overall Rank | 502 (0, 99.94) | 51.71 (26.83) | 0.113 (0.011) |
| ***Region*** |  |  |  |
| Midwest | 69 (−2.87, 4) | 0.35 (1.32) |  |
| Northeast | 130 (−3.93, 7.91) | 1.78 (2.22) |  |
| South | 199 (−6.25, 8.59) | 1.62 (2.07) |  |
| West | 104 (−2.89, 8.75) | 0.87 (1.99) |  |
| ***Locale*** |  |  |  |
| City | 167 (−3.93, 8.75) | 1.65 (2.3) |  |
| Rural | 112 (−6.25, 7.5) | 1.04 (1.92) |  |
| Suburb | 169 (−2.93, 8.59) | 1.3 (1.96) |  |
| Town | 54 (−2.87, 7.30) | 1.05 (1.84) |  |

Table 27 reviews summary statistics and t-test results for the 10 school district policies with the change in the school-level case rate as the outcome variable. *No policy* reflects the change in case rate among schools without district guidance on prevention strategy implementation in the fall of 2021, whereas *Has policy* indicates the change in case rate among schools with district guidance on prevention strategy implementation. *Difference in means* provides the mean difference between groups, calculated by subtracting the policy from the no policy mean. Seven of the strategies were significantly associated with larger increases in case rates between semesters, including staying home when sick (mean difference = .44; p-value = 0.04), quarantining (mean difference = .68; p-value = .01), cleaning (mean difference = .60; p-value = .03), and upkeep of HVAC systems (mean difference = 0.53; p-value = 0.04).

Table 27. Summary statistics and t-test results of COVID-19 prevention strategies

| Construct | n (min, max) | Overall mean (SD) | No policy | Has policy | Difference in means  (p-value) |
| --- | --- | --- | --- | --- | --- |
| Vaccination offered | 502 (0, 1) | 0.07 (0.26) | 1.34 | 1.29 | 0.04 (0.904) |
| Universal masking requirements | 502 (0, 1) | 0.12 (0.32) | 1.29 | 1.67 | 0.38 (0.213) |
| Physical distancing | 502 (0, 1) | 0.13 (0.33) | 1.36 | 1.12 | -0.25 (0.357) |
| Screening and testing for students | 502 (0, 1) | 0.12 (0.32) | 1.29 | 1.63 | 0.332 (0.293) |
| Staying home when sick | 502 (0, 1) | 0.29 (0.46) | 1.20 | 1.64 | 0.439 (0.035) |
| Contact tracing | 502 (0, 1) | 0.21 (0.40) | 1.31 | 1.41 | 0.097 (0.683) |
| Quarantining | 502 (0, 1) | 0.20 (0.40) | 1.20 | 1.87 | 0.679 (0.006) |
| Cleaning | 502 (0, 1) | 0.17 (0.38) | 1.23 | 1.83 | 0.599 (0.027) |
| HEPA filters | 502 (0, 1) | 0.02 (0.15) | 1.34 | 1.34 | 0.417 (0.31) |
| HVAC systems | 502 (0, 1) | 0.19 (0.39) | 1.23 | 1.23 | 0.529 (0.035) |

Table 28 shows results from the first set of multilevel models that were run individually for each strategy. These models adjusted for the following covariates: percent student body Asian, percent student body Black or African American, percent student body two or more races, percent student body White, percent student body free or reduced price meals, and county-level indicators including change in COVID-19 case rate and SVI Overall Rank. As shown in table three, three strategies were statistically significantly associated with increased changes in case rates, including quarantining (coefficient = .52; p-value = .02), staying home when sick (coefficient = .42; p-value = .03), and cleaning (coefficient = .46; p-value = .05). One strategy, upkeep of HVAC systems (coefficient = .42; p-value = .06), demonstrated a marginal association. As such, these strategies were selected for calculation of the cumulative indices, and are reviewed below in table five.

**Table 28. Results of multilevel models for each individual strategy accounting for covariatesa**

| **Strategy** | **Coefficient (95% interval)** | **p-value** |
| --- | --- | --- |
| Vaccination offered | −0.02 (−0.69, 0.61) | 0.94 |
| Universal masking requirements | 0.22 (−0.28, 0.76) | 0.42 |
| Physical distancing | −0.04 (−0.52, 0.48) | 0.89 |
| Screening and testing for studentsc | 0.35 (−0.22, 0.89) | 0.22 |
| Staying home when sickb | 0.42 (0.03, 0.83) | 0.03 |
| Contact tracing | 0.18 (−0.24, 0.60) | 0.41 |
| Quarantiningb | 0.52 (−0.72, 0.92) | 0.02 |
| Cleaning b | 0.46 (0.01, 0.92) | 0.05 |
| HEPA filters | 0.03 (−1.15, 1.11) | 0.95 |
| HVAC systemsb | 0.42 (−0.05, 0.82) | 0.06 |

aFull list of covariates can be found in Table 2

bincluded in 4-strategy index; c included in 5-strategy index

Table 29 shows multilevel model results from including all strategies as predictors in one model. One strategy, physical distancing, was significantly associated with a smaller increase in case rates (coefficient = -.72; p-value = 0.04).

Table 29. Results of multilevel model including all strategies

| Strategy | Coefficient (95% interval) | p-value |
| --- | --- | --- |
| Intercept | 1.04 (0.33, 1.82) | 0.07 |
| Vaccination offered | -.28 (−1.12, 0.51) | 0.49 |
| Universal masking requirements | -0.26 (−1.07, 0.42) | 0.49 |
| Physical distancing | −0.72 (−1.33, -0.03) | 0.04 |
| Screening and testing for students | 0.07 (−0.76, 0.82) | 0.87 |
| Staying home when sick | 0.24 (−0.40, 0.84) | 0.46 |
| Contact tracing | -0.19 (−0.77, 0.40) | 0.53 |
| Quarantining | 0.58 (−0.07, 1.30) | 0.11 |
| Cleaning | 0.32 (−0.25, 0.95) | 0.30 |
| HEPA filters | 0.05 (−1.29, 1.41) | 0.94 |
| HVAC systems | 0.32 (−0.36, 1.00) | 0.33 |
| Percent two or more races | 0.13 (-0.12, 0.34) | 0.31 |
| Percent Asian | 0.05 (−0.22, 0.31) | 0.70 |
| Percent White | −0.23 (−1.50, 0.60) | 0.58 |
| Percent free and reduced lunch | −0.20 (−0.51, 0.08) | 0.18 |
| SVI Overall Rank | 0.10 (−0.12, 0.30) | 0.39 |
| Percent Black or African American | −0.27 (−0.80, 0.28) | 0.34 |
| Change in county COVID-19 case rate | 0.18 (−0.12, 0.49) | 0.26 |

Results from the multilevel models, including the cumulative indices as the predictors of interest, are depicted in Table 30. The first set of results correspond with the cumulative index calculated using strategies having a marginal association with case rates (p-value < 0.10; four strategies) and the second with the sum of strategies that had p-values < 0.30 (five total). Both indices were treated as ranked ordinal, with a minimum of zero (baseline). Finally, a third column of results are presented to further explore differences between the two cumulative indices.

Schools located in districts with prevention guidance on all four strategies, including staying home when sick, quarantining, cleaning, and upkeep of HVAC systems, had a significantly higher increase in case rates between fall and spring (0.72; 0.12 - 1.41). When offering screening and testing for students was added to the cumulative index, districts with prevention guidance on all five strategies had a significantly increased change in case rates (1.09; 0.24 - 1.88).

To further compare the cumulative indices for selection of the best combination, we developed mutually exclusive indicators for having all four or all five strategies. These results are presented in the last column of table 5. Having the first three strategies was no longer statistically significant (-0.10; -1.18 – 0.98), but the indicator for having all five strategies was significantly associated with a larger change in case rate (1.09; 0.18 – 1.94). Estimates which are comparable to those are presented in the second column. These results further suggest the utility of the five-category indicators model.

Table 30. Results of multilevel model with cumulative index of marginally significant strategies

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Cumulative number strategies | Coefficient (95% CI) | p-value | Coefficient(95% CI) | p-value | Coefficient(95% CI) | p-value |
| Cutoff |  | 0.1 |  | 0.2 |  | Comparison |
| 1 | -0.21 (-0.82, 0.42) | 0.51 | -0.19 (-0.81, 0.46) | 0.55 |  |  |
| 2 | 0.27 (-0.34, 0.97) | 0.42 | 0.17(-0.62, 0.82) | 0.63 |  |  |
| 3 | 0.47 (-0.14, 1.11) | 0.13 | 0.61 (-0.14, 1.22) | 0.07 | -0.10 (-1.18, 0.98) | 0.84 |
| 4 | 0.72 (0.09, 1.35) | 0.03 | -0.11 (-0.90, 0.63) | 0.77 |  |  |
| 5 |  |  | 1.09 (0.24, 1.88) | 0.01 | 1.09 (0.18, 1.94) | 0.01 |

#### Analytic Limitations

3policy documents that included guidance onlargerfurther studying Eten

#### Summary

Despite these limitations and that district policies were associated with increased COVID-19 incidence, this analysis suggests the importance of district-level guidance as a mechanism for inducing prevention strategies among schools. Specifically, we found that districts with publicly available school district policy guidance on four strategies, upkeep (encouraging replacing, upgrading, maintaining, or inspecting) of HVAC systems, quarantining, encouraging that students stay home when sick, and cleaning schools, may have increased COVID-19 spread. These findings are further supported by the effect of combining or layering a combination of five strategies, including upkeep of HVAC systems, cleaning schools, encouraging that students stay home when sick, requiring quarantining, and offering screening and testing for students, indicated by a statistically significant association between changes in school-level COVID-19 case rates and the five-strategy cumulative index. The combined five-strategy indicator also had a stronger association with change in case rates when compared with the three-strategy indicator. the positive association between clear for and changes in COVID-19 may also suggest that districts adopted strategies as COVID-19 in the school and community increased in prominence