JAMA Network Open

Original Investigation | Infectious Diseases

COVID-19 Incidence and Age Eligibility for Elementary School

Eve Lin, BS; Alyssa Bilinski, PhD; Philip A. Collender, MPH; Vivian Lee, BS; Sohil R. Sud, MD; Tomás M. León, PhD; Lauren A. White, PhD; Justin V. Remais, PhD; Jennifer R. Head, PhD, MPH

Abstract

IMPORTANCE Understanding the role of school attendance on transmission of SARS-CoV-2 among children is of importance for responding to future epidemics. Estimating discontinuities in outcomes by age of eligibility for school attendance has been used to examine associations between school attendance and a variety of outcomes, but has yet to be applied to describe associations between school attendance and communicable disease transmission.

OBJECTIVE To estimate the association between eligibility for elementary school and COVID-19 incidence.

DESIGN, SETTING, AND PARTICIPANTS This case series used data on all pediatric COVID-19 cases reported to California's disease surveillance system between May 16, 2020, and December 15, 2022, among children within 24 months of the age threshold for school eligibility.

EXPOSURE Birthdate before or after the age threshold for elementary school eligibility during periods when school was remote vs in person.

MAIN OUTCOMES AND MEASURES COVID-19 cases and hospitalizations.

RESULTS Between May 16, 2020, and December 15, 2022, there were 688 278 cases of COVID-19 (348 957 cases [50.7%] among boys) and 1423 hospitalizations among children who turned 5 years within 24 months of September 1 of the school year when their infection occurred. The mean (SD) age of the study sample was 5.0 (1.3) years. After adjusting for higher rates of testing in schooled populations, the estimated pooled incidence rate ratio among kindergarten-eligible individuals (eg, those born just before the age threshold for school eligibility) compared with those born just after the eligibility threshold for in-person fall 2021 semester was 1.52 (95% CI, 1.36-1.68), for in-person spring 2022 semester was 1.26 (95% CI, 1.15-1.39), and for in-person fall 2022 semester was 1.19 (95% CI, 1.03-1.38). Reported incidence rates among school-eligible children remained higher during the month-long winter 2021-2022 school break but were lower during the longer summer break that followed. The findings were unable to establish whether associations between school eligibility and COVID-19 incidence were based on in-school vs out-of-school routes (eg, classrooms vs school buses). The study lacked power to detect associations between school attendance and hospitalization. Results were robust to functional form. A simulation study was conducted to demonstrate bias associated with nonadjustment for differential case acquisition by exposure status.

CONCLUSIONS AND RELEVANCE In this case series of children in California, the magnitude of the association between school eligibility and COVID-19 incidence decreased over time and was generally lower than other published associations between out-of-school child social interactions and COVID-19 incidence. This regression discontinuity design approach could be adapted to other geographies and/or disease systems to assess associations between schooling and disease transmission.

JAMA Network Open. 2024;7(11):e2444836. doi:10.1001/jamanetworkopen.2024.44836

Key Points

Question Was eligibility for elementary school associated with COVID-19 infection in California?

Finding In this case series study, using regression discontinuity methods that adjusted for differential testing rates in schooled populations, higher incidences of COVID-19 were found among California children eligible for kindergarten compared with children born just after the age threshold for school eligibility during in-person semesters: fall 2021 (51.5% higher), spring 2021 (26.3% higher), and fall 2022 (19.1% higher). No associations were found between school eligibility and hospitalization.

Meaning This study suggests that associations between school eligibility and COVID-19 incidence decreased over time and were generally smaller than published associations between out-of-school gatherings and incidence.

+ Multimedia

+ Supplemental content

Author affiliations and article information are listed at the end of this article.

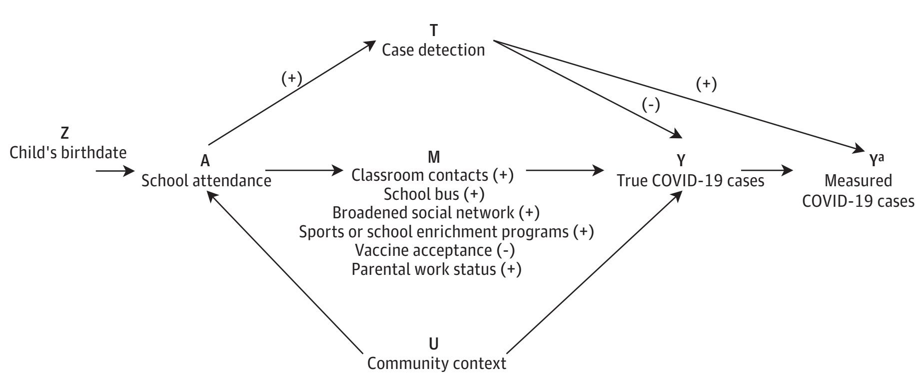
Introduction

Although the association of school attendance with the transmission of SARS-CoV-2 among children emerged as one of the most important questions of the pandemic, it proved difficult to answer.[1] Ecological studies examining changes in case growth rates or reproduction number before and after school closures were limited in their ability to disentangle the effect of school interventions from concurrent policies (eg, workplace closures, social gathering bans).[2,3] Prospective studies observing students in classroom settings provided some evidence that COVID-19 incidence among students mimicked rates in the general community, while offering methodological benefits such as controlled testing and examination of multiple end points.[4-8] However, these studies can be costly to conduct and are too time consuming during early stages of a pandemic, when decisions must be made quickly. Furthermore, research suggests that the effect of schooling on COVID-19 incidence depends on local mitigation efforts[1] and vaccination rates,[9] limiting generalizability of findings from focal cohort studies.

Given the substantial tolls of school closures on childhood education[10,11] and mental health,[12,13] as well as the exacerbation of social and racial and ethnic inequalities in these outcomes,[10,14,15] methods are needed to rapidly and robustly quantify the association of school attendance with transmission rates of newly emerged pathogens or variants, including SARS-CoV-2 variants or novel influenza strains. The age threshold for school eligibility (typically whether a child's fifth birthday falls before September 1)[16] is a promising instrumental variable for school exposures that can be used within a regression discontinuity design (RDD) to yield insights into outcomes associated with schooling. Because children born immediately before the age threshold for school eligibility are assumed to be similar to those born immediately after, any discontinuous changes in outcomes as age crosses the threshold can be interpreted as being associated, at least in part, with school attendance.[17,18] Use of age-based school eligibility data avoids the extensive data collection required by cohort studies or the experimentation in policy necessary for ecological studies.

Regression discontinuity designs have been used to examine associations between school attendance and crime initiation,[19] adult earnings,[20] adolescent obesity,[21] and maternal anxiety,[22] but, to our knowledge, have not been applied to understand the association between school attendance and infectious disease outcomes.[23] School attendance may be associated with disease transmission via both in-school contacts and out-of-school consequences of schooling (eg, riding buses, parents returning to work; **Figure 1**). Although routine collection of data on reportable diseases enhances the feasibility of using RDDs to study associations between school attendance and infectious diseases outcomes, measurement error of the outcome via imperfect case ascertainment may bias effect estimates if case ascertainment is differential by exposure status. This concern is salient for COVID-19, as in-school symptomatic and asymptomatic testing programs were a component of many COVID-19 prevention plans (Figure 1).

Figure 1. Directed Acyclic Graph Representing Associations Between Child Birthdate, School Attendance, and COVID-19 Cases



(+) Indicates a positive association; (−) indicates a negative association.

[a] Signifies the measured version vs the true version.

Here, we explore RDD as a means of estimating the association between school attendance and an infectious disease in the presence of differential case ascertainment by comparing COVID-19 incidence and hospitalizations among children born just after the age threshold for kindergarten eligibility with those born just before. California was chosen as the setting for this study because it has the largest public school enrollment in the US[24] and is the most socioeconomically diverse state.[25]

Methods

Data Source

We obtained information on all pediatric COVID-19 cases reported in California between May 16, 2020, and December 15, 2022, from the California COVID-19 Reporting System. COVID-19 cases were confirmed using a positive nucleic acid amplification test. Each record contained information on the patient's census tract of residence, age in months, and hospitalization status. Because the underlying data were collected for public health surveillance and deidentified, the secondary research described herein is exempt from institutional review board review and consent according to the Common Rule (45 CFR §46). We followed the reporting guidelines for case series studies.

We defined school periods as fall semester (August 15-December 14), winter break (December 15-January 14), spring semester (January 15-May 15), and summer break (May 16-August 14) (eFigure 5 and eTable 3 in Supplement 1).[26,27] We examined 2 semesters of remote instruction (2020-2021 academic year) and 3 semesters of in-person instruction (fall and spring of the 2021-2022 academic year and fall of the 2022-2023 academic year).

For each academic year being analyzed, we assessed the subsample of individuals within the surveillance record who would fall within a given bandwidth, h, of the age threshold for school eligibility (eTable 3 in Supplement 1). In California, as in most states, children must be 5 years of age by September 1 to enter kindergarten. Within each county and school period (ie, fall and spring semesters, summer and winter breaks), we calculated the number of reported cases, stratified by child's birth month. Hospitalization data were summarized similarly to case data, except we summed cases to the state level, rather than county level, due to the relative rarity of hospitalizations.

To obtain population denominators, we used data on births per county and month that gave rise to the underlying population.[28] For instance, for the 2020-2021 academic year, we obtained county-level data on the number of children born each month in the bandwidth around September 1, 2015 (eTable 3 in Supplement 1).

Statistical Analysis

Regression Discontinuity Design

We used a sharp RDD to estimate the association between school attendance and reported COVID-19 cases and hospitalizations among young children, assuming that age at September 1 is associated with whether or not an individual attends school that year.

For the sample of data with $-h < x_i < h$, where h is bandwidth and x_i is the difference in age (in months) from the school eligibility threshold on September 1 of the relevant period, grouped into 1-month bins, we ran Poisson regressions with the form:
$\log [E(Y_i, | Z_i, x_i)] = \log(b_i) + \alpha + \tau Z_i + f(x_i) + g(x_i Z_i)$, where Y_i is the number of COVID-19 cases or hospitalizations among children in age bin i; b_i is the number of births of children in age bin i, approximating population size; Z_i is a binary indicator for school eligibility in the current period (ie, whether, $x_i > 0$); f is a function associating age with the outcome among noneligible children; g is a function describing the modification of the association between age and outcome among eligible children; α is an intercept term; and τ, the target parameter, is the coefficient describing the marginal association of eligibility for school attendance with the outcome. Exponentiating τ yields the incidence rate ratio (IRR) comparing COVID-19 among children born just before the age threshold for school eligibility with children born just after the threshold. Because our primary exposure is age

eligibility for school attendance rather than attendance, the estimated IRR is considered an intention-to-treat effect estimate.[29] We used 95% CIs to indicate statistical significance.

To contend with bias from differential case ascertainment, we weighted cases in the school-ineligible strata to approximate the number of cases that would have been observed in the school-ineligible age groups given equal testing effort (see next section). We did not apply weights for analyses with hospitalization as the outcome, as ascertainment of severe cases is more likely to be similar between populations.

Children born between September 2 and December 2 may have attended California's transitional kindergarten (TK) program.[30] Although TK classes are smaller than kindergarten classes, TK may result in similar exposures as elementary school attendance. In main analyses, we removed individuals eligible for TK and we included them in sensitivity analyses.

For cases, we fit separate models for 46 of California's 58 counties, excluding 12 counties with few cases and births reported. We pooled IRRs using a meta-analysis approach that has been used to combine effect estimates across multiple locations while examining effect heterogeneity (R mvmeta package; eAppendix in Supplement 1).[31] All analyses were conducted in R, version 4.2.0 (R Project for Statistical Computing).[32]

Adjustment for Differential Case Ascertainment

As case detection was a feature of California's reopening plans for kindergarten through grade 12, testing rates were higher among school-aged children compared with younger children (eFigure 6 in Supplement 1). To contend with bias from differential case ascertainment, we conducted a simulation study to identify a weighting factor, that, when used to upweight cases in the school-ineligible strata, generated an estimated IRR that most closely approximated the true IRR (eAppendix in Supplement 1). Briefly, we used a susceptible–exposed–infectious (asymptomatic/symptomatic)–recovered model to simulate both true and observed cases according to the hypothesized data-generating mechanism and expected transmission parameters for the study population (eFigures 1 and 2 and eTables 1 and 2 in Supplement 1). We then fit a regression discontinuity model to the simulated true and observed cases to calculate true and observed IRRs. Upweighting cases in the school-ineligible strata by the square root of the county's testing ratio for the 5- to 10-year age group vs the 0- to 4-year age group best allowed the observed IRR to approximate the true IRR (eAppendix and eFigures 3 and 4 in Supplement 1), a result consistent with adjustment functions arrived at in other studies using alternative means of estimation.[33] To provide another bound on estimated IRRs, we upweighted cases in the school-ineligible strata by the county's testing ratio for the 5- to 10-year age group vs the 0- to 4-year age group. This approach is considered conservative because testing was conducted as a screening tool among asymptomatic populations, so each test administered had a lower probability of a positive result.

Sensitivity Analyses, Negative Controls, and Power Simulation

To assess the robustness of results to model specification, we ran analyses varying bandwidth h from 8 to 24 months and testing 3 different functional bases (linear, quadratic, locally linear, or locally estimated scatterplot smoothing [loess]) for the association of age x_i with COVID-19 outcomes (eAppendix in Supplement 1). We compared model fit using the Akaike information criterion (AIC).[34,35]

As a negative control to identify the presence of unmeasured confounders, we examined whether there were discontinuities in COVID-19 incidence at the age-eligibility threshold during periods when we should not expect an increase in cases, including semesters when school was remote and summer and winter breaks prior to reopening of in-person instruction.

Insufficient power can be a limitation of regression discontinuity analyses,[36] a concern salient for our hospitalization analysis given the relative rarity of the outcome. We performed a power simulation study to examine the plausibility that associations of school eligibility with hospitalization were not detected due to rarity of the hospitalization outcome (eAppendix in Supplement 1).[37]

Results

Population Characteristics

Between May 16, 2020, and December 15, 2022, there were 688 278 cases of COVID-19 (348 957 [50.7%] among boys and 339 321 [49.3%] among girls) and 1423 hospitalizations among children who turned 5 years within 24 months of September 1 of the school year when their infection occurred. The mean (SD) age of the study sample was 5.0 (1.3) years. The number of cases reported during in-person semesters ranged from 28 088 to 212 093 cases and 121 to 318 hospitalizations (eTable 3 in Supplement 1).

Association Between Elementary School Attendance and Reported COVID-19 Cases

For 36 (78.3%) of the 46 counties examined, models with linear associations between age and COVID-19 incidence had lower AICs than models using local linear regression or quadratic terms. Here, we present results from linear models fitted using a bandwidth *h* of 24 months.

Reported COVID-19 incidence was higher among children eligible for school attendance during in-person semesters in the study period (**Table**; **Figure 2** shows unweighted visual discontinuities for Los Angeles County; eFigures 7 and 8 in Supplement 1 show visual discontinuities using different functional forms). Using our best correction for differential case ascertainment, the pooled estimate of the school eligibility IRR for COVID-19 during the fall 2021 semester was 1.52 (95% CI, 1.36-1.68) (Table and **Figure 3**A), meaning that age-adjusted incidence rates were 51.5% (95% CI, 36.3%-68.5%) higher among children eligible for elementary school (eg, those born just before the age threshold for school eligibility) compared with ineligible children. The IRR for the winter break that followed the fall 2021 semester was 1.33 (95% CI, 1.19-1.49). Incidence rate ratios for the spring 2022 semester were 1.26 (95% CI, 1.15-1.39) and the fall 2022 semester was 1.19 (95% CI, 1.03-1.38). Age-adjusted incidence rates for the spring 2022 semester were 26.3% (95% CI, 14.9%-38.8%) higher and for the fall 2022 semester were 19.1% (95% CI, 3.0%-37.3%) higher among children eligible for elementary school compared with ineligible children. eFigures 9 to 11 in Supplement 1 show county-specific IRRs.

Including individuals eligible for TK resulted in effect estimates closer to the null (Table; eFigures 12 and 13 in Supplement 1). Using the more conservative weighting factor resulted in effect estimates closer to the null, which were significant only for the 2021-2022 academic year (fall 2021 IRR, 1.36 [95% CI, 1.20-1.53]; spring 2022 IRR, 1.14 [95% CI, 1.03-1.28]) (eTable 4 and eFigure 12 in Supplement 1). Not adjusting for differential testing resulted in effect estimates further from the null. This finding aligns with results from our simulation study, which demonstrates deviation between

Table. Incidence Rate Ratios Comparing the Incidence of COVID-19 Among Children Born Just Before the Threshold for Elementary School Attendance (September 1) Compared With Just After

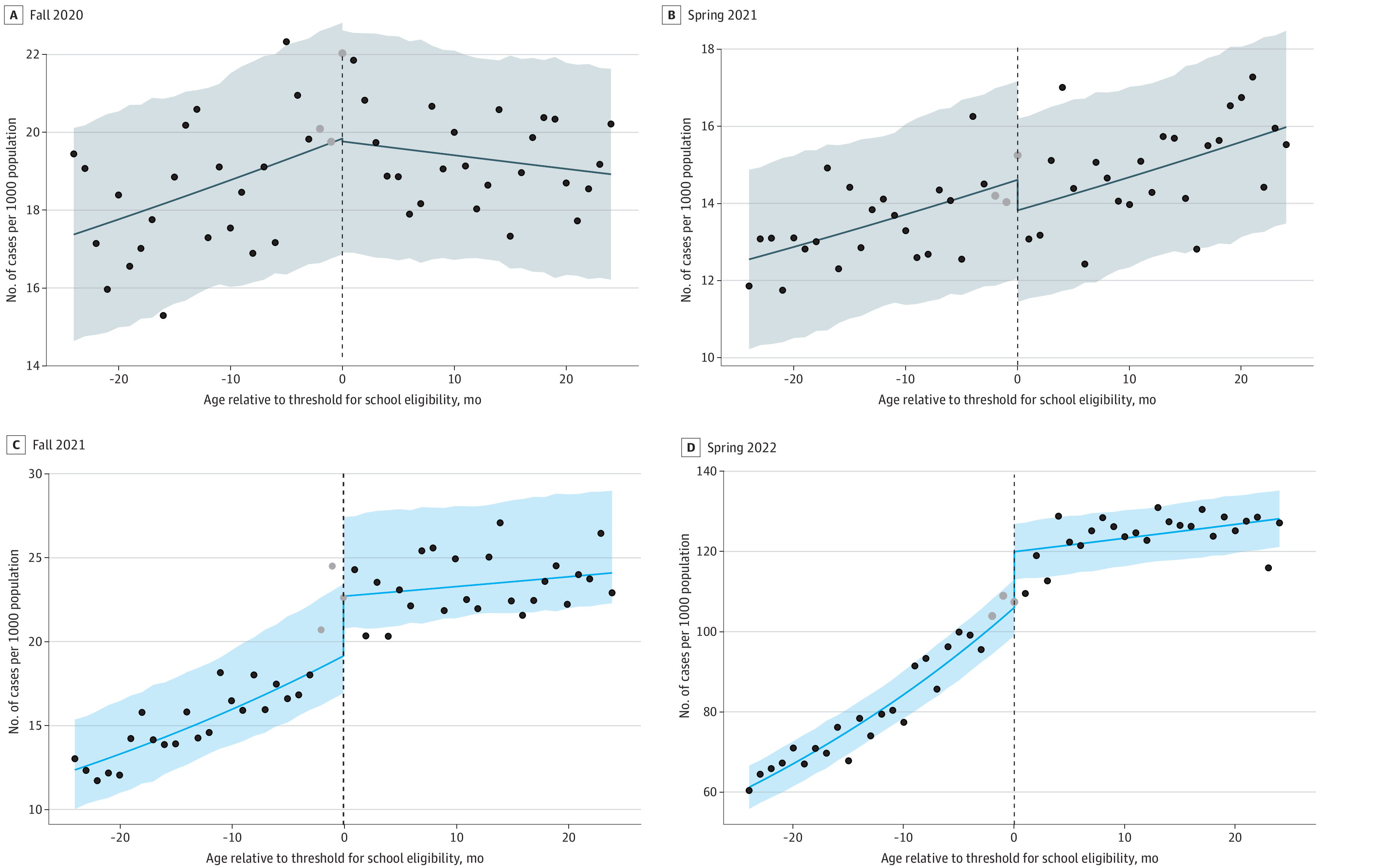
School period	Incidence rate ratio (95% CI)	
	Excluding TK-eligible children[a]	Including TK-eligible children[a]
Prior to reopening of schools for in-person instruction		
Summer 2020	1.01 (0.86-1.19)	1.00 (0.96-1.16)
Fall 2020	0.89 (0.79-1.01)	0.95 (0.85-1.06)
Winter 2020	0.93 (0.80-1.08)	0.96 (0.84-1.10)
Spring 2021	1.03 (0.92-1.15)	1.03 (0.93-1.14)
Summer 2021	0.88 (0.77-1.00)	0.94 (0.84-1.06)
After reopening of schools for in-person instruction		
Fall 2021[b]	1.52 (1.36-1.68)	1.23 (1.13-1.34)
Winter 2021	1.33 (1.19-1.49)	1.24 (1.12-1.37)
Spring 2022[b]	1.26 (1.15-1.39)	1.11 (1.02-1.19)
Summer 2022	0.87 (0.76-0.99)	0.91 (0.80-1.03)
Fall 2022[b]	1.19 (1.03-1.38)	1.06 (0.94-1.21)

Abbreviation: TK, transitional kindergarten.

[a] Children born between September 2 and December 1 are eligible for TK, so we conducted analyses both excluding and including children born from September 2 to December 1.

[b] In-person semester included in the study period.

Figure 2. COVID-19 Incidence as a Function of Children's Age Relative to the September 1 Age Threshold for School Eligibility



Ages to the right of the threshold indicate that the child is age eligible to attend elementary school (kindergarten through grade 5). Model fits are shown for the fall 2020 (A) and spring 2021 (B) semesters when school was remote and during the fall 2021 (C) and spring 2022 (D) semesters when school was in-person. Dots indicate observed data, the gray dots indicate data for the children who were eligible for transitional kindergarten and excluded from main analyses, the lines indicate model fit, and the shaded regions indicate 95% CIs. For this example, weighting to adjust for testing biases is not performed. Plots shown are selected from Los Angeles County, which contains the largest school district in California. Models shown assume a linear association between age and incidence, use a bandwidth of 24 months, and exclude children eligible for transitional kindergarten. Fits from other functional forms are shown in eFigures 7 and 8 in Supplement 1.

the true and observed IRR when outcome measurement is associated with exposure (eAppendix in Supplement 1).

No significant differences in COVID-19 case incidence were observed during academic periods prior to the reopening of schools for in-person instruction (Table; Figure 3A). Reported incidence remained higher among school-eligible children during the month-long winter 2021-2022 break that followed the in-person fall 2021 semester (IRR, 1.33 [95% CI, 1.19-1.49]), and was lower during the longer summer 2022 break that followed in-person 2021-2022 academic year (IRR, 0.87 [95% CI, 0.76-0.99]). Results were robust to functional forms (**Figure 4**).

Of county-level variables examined, only population explained heterogeneity in school eligibility IRR across counties in meta-regression (eTable 5 in Supplement 1). Higher county population was associated with lower IRRs across all 3 in-person semesters, suggesting that associations between schooling and COVID-19 incidence were weaker among more populous counties. Counties with a higher proportion of the population who reported never wearing a mask and with a higher percentage of single-parent families had higher school eligibility IRRs across all 3 in-person semesters, but this association was not significant at the 95% confidence level.

Association Between Elementary School Attendance and Reported COVID-19 Hospitalizations

There were no discernible increases in reported age-adjusted COVID-19 hospitalizations rates among children who were born before the school attendance age threshold for any period examined (eFigures 14-16 in Supplement 1) and alternative model parameterizations (eFigure 17 in Supplement 1). This finding could have been due to the rarity of hospitalization for COVID-19 among children (eFigure 18 and eAppendix in Supplement 1).

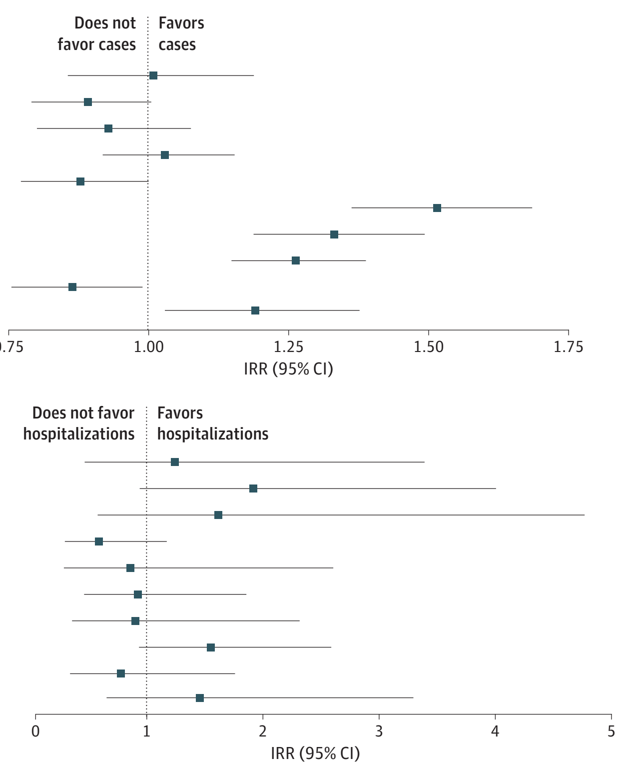
Figure 3. Incidence of COVID-19 Cases and Hospitalizations

A Cases

School period	IRR (95% CI)
Summer 2020	1.01 (0.86-1.19)
Fall 2020	0.89 (0.79-1.01)
Winter 2020	0.93 (0.80-1.08)
Spring 2021	1.03 (0.92-1.15)
Summer 2021	0.88 (0.77-1.00)
Fall 2021[a]	1.52 (1.36-1.68)
Winter 2021	1.33 (1.19-1.49)
Spring 2022[a]	1.26 (1.15-1.39)
Summer 2022	0.87 (0.76-0.99)
Fall 2022[a]	1.19 (1.03-1.38)

B Hospitalizations

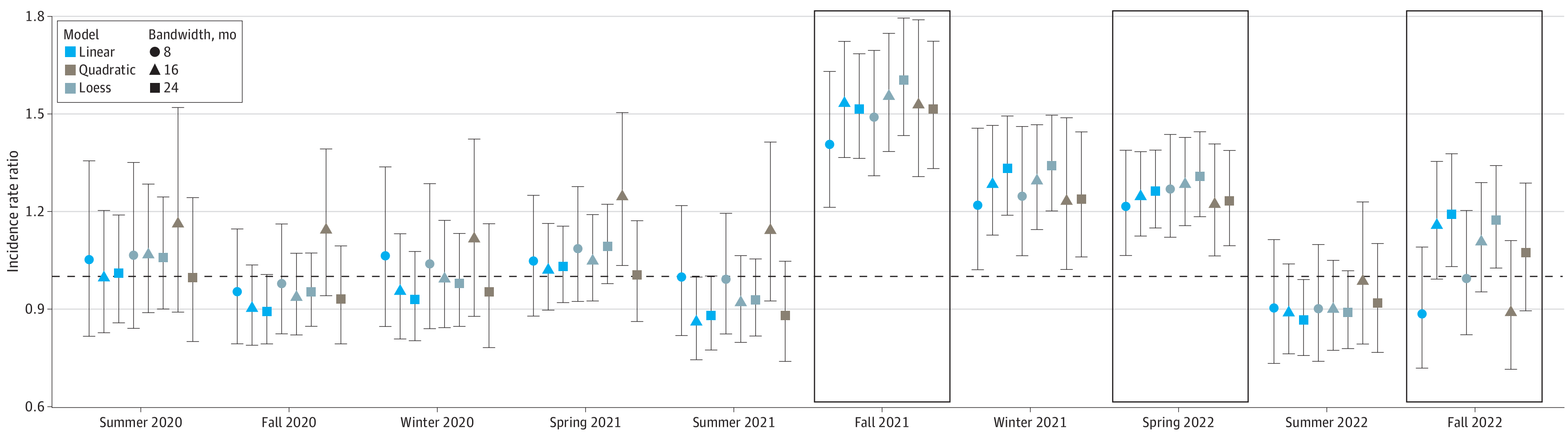
School period	IRR (95% CI)
Summer 2020	1.24 (0.47-3.39)
Fall 2020	1.92 (0.94-4.00)
Winter 2020	1.62 (0.58-4.76)
Spring 2021	0.59 (0.30-1.17)
Summer 2021	0.86 (0.29-2.60)
Fall 2021[a]	0.92 (0.46-1.86)
Winter 2021	0.90 (0.36-2.32)
Spring 2022[a]	1.55 (0.93-2.59)
Summer 2022	0.78 (0.34-1.76)
Fall 2022[a]	1.46 (0.66-3.29)



Incidence rate ratios (IRRs) and 95% CIs (indicated by horizontal lines) representing incidence of COVID-19 cases (pooled across counties) (A) and hospitalizations (statewide) (B) among children born just before the threshold for elementary school attendance (kindergarten through grade 5) compared with those born just after. Models shown assume a linear association between age and outcome, use a bandwidth of 24 months, exclude children eligible for transitional kindergarten (TK), and adjust for bias due to unequal case ascertainment. eFigure 13 in Supplement 1 shows results including individuals eligible for TK.

[a] Periods when schools were open for in-person instruction.

Figure 4. Incidence of COVID-19 Cases and Hospitalizations



Incidence rate ratios and 95% CIs (indicated by vertical lines) representing incidence of COVID-19 cases among children born just before the threshold for elementary school attendance (kindergarten through grade 5) compared with those born just after. Incidence rate ratios are displayed for various model parametrizations (eg, local linear regression [locally estimated scatterplot smoothing (Loess)], linear association between age and incidence, and quadratic association between age and incidence) and bandwidths. In-person semesters are outlined in black boxes. Quadratic models with 8-month bandwidths are not shown due to large CIs. Graphical results for hospitalizations are displayed in eFigure 18 in Supplement 1.

Discussion

Using a regression discontinuity approach and adjusting for greater case ascertainment among school-aged populations, we estimated increases in COVID-19 incidence of 51.5% during the fall 2021 semester, 26.3% during the spring 2022 semester, and 19.1% during the fall 2022 semester among children born just before the age threshold for kindergarten eligibility in California compared with children born just after during the first 3 semesters when schools were open for in-person instruction. Overall, the estimated effect sizes for the 3 in-person semesters are similar in magnitude to or smaller than the estimated effect size of child social gatherings (increase of 31% in COVID-19 incidence after birthdays)[38] and smaller than the effect size associated with large gatherings among adults.[39]

Our findings on associations between in-person school attendance and COVID-19 case incidence align with prior research. Although reported outbreaks among classroom and daycare settings suggest that these locations hold potential for spreading disease,[40-44] evidence suggests that precautions were able to effectively mitigate large increases in transmission.[4,5,11] Prior work also demonstrates slight increases in incidence among teachers[45] and household members of students, which can be minimized by in-school prevention measures.[1] Model-based[9,46] and empirical studies have found limited associations between schooling and severe pediatric COVID-19,[11,47,48] which, as in our study, may be associated with the relative rarity of severe outcomes among children. We found that COVID-19 incidence increases linearly with age, controlling for school eligibility (Figure 2), a trend that has been observed elsewhere.[49] When considering the evidence associating school closures with learning loss,[10] adverse mental health,[12,13] and widening disparities,[14,15] our findings and those of others support the use of within-school cautionary measures as much as possible over school closures.[50]

The association between school eligibility and COVID-19 incidence decreased with each subsequent in-person semester. In most model parameterizations, school eligibility appeared to be associated with protection against COVID-19 incidence during the summer after the first in-person school year that followed widespread school closures. These findings may suggest a protective association of natural immunity due to higher infection rates and/or greater adoption of vaccines, which became available in late October of the fall 2021 semester.[51]

In meta-analyses, we found that more populated counties had weaker associations between in-person schooling and COVID-19 incidence. This finding could reflect the fact that major school districts in populated counties, including Los Angeles, San Diego, San Francisco, and Sacramento, adhered to stricter within-school mitigation measures, including longer mask mandates.[52] The percentage of the population who reported never wearing a mask was associated with stronger associations between in-person schooling and COVID-19 incidence, although not significantly so at the 95% confidence level. More populous counties also had a higher cumulative incidence of COVID-19 at the time of school reopening, suggesting higher immunity of schooled children and/or increased transmission among the comparison group of school-ineligible children.

Limitations

This study has some limitations. We cannot identify whether higher COVID-19 incidence among children born before vs after the school attendance threshold are due to in-school or out-of-school exposures that are associated with school attendance, such as bus ridership or sports programs (Figure 1). Prior work suggests that such community settings pose a higher risk than classrooms of infection.[53] Second, the comparison group in this study—children younger than the threshold for school attendance—has heterogeneous contact patterns that could vary between staying at home or being placed in daycare. We treated our RDD as a sharp design, assuming all children eligible for school will attend school, and all children not eligible for school will not attend school. However, children may be homeschooled, held back, or sent to school early. The resulting bias from exposure misclassification is expected to be toward the null.

We modeled surveillance data, so jumps in discontinuities at the threshold may reflect increased testing, a bias we attempted to adjust for in analyses. The estimated IRR for the fall 2021 semester (1.52) is similar to that for the winter break that followed (1.33), providing some validity for our bias-adjusted estimate, as school-associated disparities in testing volume should have been smaller over the winter break period, yet elevated incidence associated with school attendance could have lingered due to latency periods and secondary transmission. The testing data we used to estimate adjustment factors did not account for differential testing over time, which would have permitted use of alternative methods that have been published that compare incidences adjusted for differential testing.[54] Nevertheless, our weighting factor is consistent with previous literature that uses convex functions of testing effort to adjust ascertained cases upward.[33,54] Here, we estimate function to be a square root function, which is consistent with that estimated by other study.[33] The true IRRs may thus be closer to the null if the full extent of disparities in testing volume were dampened by inclusion of periods when school was on break. In-school testing efforts may have tapered over time, contributing, in part, to the observed decrease in IRRs over time. The challenge of differential case ascertainment by school status is likely greater for COVID-19 than for other pediatric infections, such as influenza or respiratory syncytial virus, where in-school surveillance testing is not common.

The effect estimates from this study may not be generalizable to states outside California, which held longer masking policies than most states and had vaccine mandates for teachers,[55] nor other countries that may have kept schools open while maintaining more stringent contact tracing or other within-school measures. Although we could not examine associations between school eligibility and COVID-19 incidence among potentially more vulnerable household members of schooled children, this approach could be used to investigate this association, if information on the age of household members of adult cases was known.

Conclusions

In this case series of pediatric COVID-19 cases, we estimated increases in reported COVID-19 incidence among children eligible for elementary school attendance in California compared with children ineligible for elementary school, and observed that these increases decreased over time. This study demonstrated the feasibility of regression discontinuity to investigate associations between COVID-19 and school attendance. This approach may be especially useful when direct observation and follow-up of school-aged children is not feasible. The novelty of using this approach to answer questions surrounding infectious disease incidence was both a strength of this study and motivator for future research to characterize sources of bias when using the regression discontinuity approach to study infectious disease transmission within schools.

ARTICLE INFORMATION

Accepted for Publication: September 22, 2024.

Published: November 14, 2024. doi:10.1001/jamanetworkopen.2024.44836

Corresponding Author: Jennifer R. Head, PhD, MPH, Department of Epidemiology, University of Michigan, 1415 Washington Heights, Ann Arbor, MI 48109 (jrhead@umich.edu).

Author Affiliations: College of Letters and Sciences, University of California, Berkeley (Lin, Lee); School of Public Health, Brown University, Providence, Rhode Island (Bilinski); Division of Environmental Health Sciences, University of California, Berkeley (Collender, Remais, Head); California Department of Public Health, Richmond (Sud, León, White); Department of Epidemiology, University of Michigan, Ann Arbor (Head); Institute of Global Change Biology, University of Michigan, Ann Arbor (Head).

Author Contributions: Dr Head had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

Concept and design: Bilinski, Collender, Remais, Head.

Acquisition, analysis, or interpretation of data: Lin, Lee, Sud, León, White, Head.

Drafting of the manuscript: Lin, Collender, White, Remais, Head.

Critical review of the manuscript for important intellectual content: Bilinski, Collender, Lee, Sud, León, White, Remais, Head.

Statistical analysis: Lin, Bilinski, Collender, León, White, Head.

Obtained funding: Remais, Head.

Administrative, technical, or material support: Lin, Sud, León, White, Remais.

Supervision: Remais, Head.

Conflict of Interest Disclosures: Dr Bilinski reported receiving grants from the Centers for Disease Control and Prevention through the Council of State and Territorial Epidemiologists during the conduct of the study; and grants from the National Center for HIV, Viral Hepatitis, STD, and Tuberculosis Prevention outside the submitted work. No other disclosures were reported.

Funding/Support: This project was supported by a grant from the California Department of Public Health through the UC Health & California Department of Public Health COVID Modeling Consortium. Dr Head was funded by grant K01AI173529 from the National Institutes of Health.

Role of the Funder/Sponsor: The funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

Data Sharing Statement: See Supplement 2.

REFERENCES

1. Lessler J, Grabowski MK, Grantz KH, et al. Household COVID-19 risk and in-person schooling. *Science*. 2021;372(6546):1092-1097. doi:10.1126/science.abh2939

2. Chernozhukov V, Kasahara H, Schrimpf P. The association of opening K-12 schools with the spread of COVID-19 in the United States: county-level panel data analysis. *Proc Natl Acad Sci U S A*. 2021;118(42):e2103420118. doi:10.1073/pnas.2103420118

3. Yang B, Huang AT, Garcia-Carreras B, et al; UFCOVID Interventions Team. Effect of specific non-pharmaceutical intervention policies on SARS-CoV-2 transmission in the counties of the United States. *Nat Commun*. 2021;12(1):3560. doi:10.1038/s41467-021-23865-8

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):136-140. doi:10.15585/mmwr.mm7004e3

5. Ismail SA, Saliba V, Lopez Bernal J, Ramsay ME, Ladhani SN. SARS-CoV-2 infection and transmission in educational settings: a prospective, cross-sectional analysis of infection clusters and outbreaks in England. *Lancet Infect Dis*. 2021;21(3):344-353. doi:10.1016/S1473-3099(20)30882-3

6. Donovan CV, Rose C, Lewis KN, et al. SARS-CoV-2 incidence in K-12 school districts with mask-required versus mask-optional policies—Arkansas, August–October 2021. *MMWR Morb Mortal Wkly Rep*. 2022;71(10):384-389. doi:10.15585/mmwr.mm7110e1

7. Macartney K, Quinn HE, Pillsbury AJ, et al; NSW COVID-19 Schools Study Team. Transmission of SARS-CoV-2 in Australian educational settings: a prospective cohort study. *Lancet Child Adolesc Health*. 2020;4(11):807-816. doi:10.1016/S2352-4642(20)30251-0

8. Heavey L, Casey G, Kelly C, Kelly D, McDarby G. No evidence of secondary transmission of COVID-19 from children attending school in Ireland, 2020. *Euro Surveill*. 2020;25(21):2000903. doi:10.2807/1560-7917.ES.2020.25.21.2000903

9. Head JR, Andrejko KL, Remais JV. Model-based assessment of SARS-CoV-2 Delta variant transmission dynamics within partially vaccinated K-12 school populations. *Lancet Reg Health Am*. 2022;5:100133. doi:10.1016/j.lana.2021.100133

10. Engzell P, Frey A, Verhagen MD. Learning loss due to school closures during the COVID-19 pandemic. *Proc Natl Acad Sci U S A*. 2021;118(17):e2022376118. doi:10.1073/pnas.2022376118

11. Munro A, Buonsenso D, González-Dambrauskas S, et al. In-person schooling is essential even during periods of high transmission of COVID-19. *BMJ Evid Based Med*. 2023;28(3):175-179. doi:10.1136/bmjebm-2023-112277

12. Hawrilenko M, Kroshus E, Tandon P, Christakis D. The association between school closures and child mental health during COVID-19. *JAMA Netw Open*. 2021;4(9):e2124092. doi:10.1001/jamanetworkopen.2021.24092

13. Lee J. Mental health effects of school closures during COVID-19. *Lancet Child Adolesc Health*. 2020;4(6): 421-421. doi:10.1016/S2352-4642(20)30109-7

14. Parolin Z, Lee EK. Large socio-economic, geographic and demographic disparities exist in exposure to school closures. *Nat Hum Behav*. 2021;5(4):522-528. doi:10.1038/s41562-021-01087-8

15. Van Lancker W, Parolin Z. COVID-19, school closures, and child poverty: a social crisis in the making. *Lancet Public Health*. 2020;5(5):e243-e244. doi:10.1016/S2468-2667(20)30084-0

16. Table 1.2. Compulsory school attendance laws, minimum and maximum age limits for required free education, by state: 2017. National Center for Education Statistics. 2017. Accessed August 9, 2023. https://nces.ed.gov/programs/statereform/tab1_2-2020.asp

17. Thistlewaite D, Campbell D. Regression-discontinuity analysis: an alternative to the ex-post facto experiment. *J Educ Psychol*. 1960;51:309-317. doi:10.1037/h0044319

18. Hahn J, Todd P, Van der Klaauw W. Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*. 2001;69:201-209. doi:10.1111/1468-0262.00183

19. Cook PJ, Kang S. Birthdays, schooling, and crime: regression-discontinuity analysis of school performance, delinquency, dropout, and crime initiation. *Am Econ J Appl Econ*. 2016;8:33-57. doi:10.1257/app.20140323

20. Angrist JD, Krueger AB. Does compulsory school attendance affect schooling and earnings? *Q J Econ*. 1991; 106:979-1014. doi:10.2307/2937954

21. Barcellos SH, Carvalho LS, Turley P. Distributional effects of education on health. *J Hum Resour*. 2023;58(4): 1273-1306. doi:10.3368/jhr.59.2.0720-11064R1

22. Takaku R, Yokoyama I. What the COVID-19 school closure left in its wake: evidence from a regression discontinuity analysis in Japan. *J Public Econ*. 2021;195:104364. doi:10.1016/j.jpubeco.2020.104364

23. Moscoe E, Bor J, Bärnighausen T. Regression discontinuity designs are underutilized in medicine, epidemiology, and public health: a review of current and best practice. *J Clin Epidemiol*. 2015;68(2):122-133. doi:10.1016/j.jclinepi.2014.06.021

24. Table 203.20. Enrollment in public elementary and secondary schools by region, state, and jurisdiction: selected years, fall 1990 through fall 2023. National Center for Education Statistics. 2023. Accessed August 14, 2023. https://nces.ed.gov/programs/digest/d13/tables/dt13_203.20.asp

25. McCann A. *Most & Least Diverse States in America*. WalletHub; 2022.

26. SFUSD calendars. San Francisco Unified School District. 2023. Accessed May 10, 2023. https://www.sfusd.edu/calendars

27. LAUSD calendars. Los Angeles Unified School District. 2023. Accessed May 10, 2023. https://achieve.lausd.net/Page/2#calendar78377/20230510/month

28. 1960-2022 Final births by month by county. California Health and Human Services Agency. Accessed February 1, 2023. https://data.chhs.ca.gov/dataset/live-birth-profiles-by-county/resource/d6c30e46-8618-407a-ba5a-bae308f86a1c

29. Oldenburg CE, Moscoe E, Bärnighausen T. Regression discontinuity for causal effect estimation in epidemiology. *Curr Epidemiol Rep*. 2016;3:233-241. doi:10.1007/s40471-016-0080-x

30. TKCalifornia. 2023. Accessed August 12, 2023. https://tkcalifornia.org

31. Gasparrini A, Armstrong B, Kenward MG. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Stat Med*. 2012;31(29):3821-3839. doi:10.1002/sim.5471

32. R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing; 2015.

33. 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. doi:10.1371/journal.pcbi.1009374

34. Akaike H. Information theory and an extension of the maximum likelihood principle. In: Petrov BN, Caski F, eds. *Proceedings of the Second International Symposium on Information Theory*. Academiai Kiado; 1973; Budapest, Hungary.

35. Symonds MRE, Moussalli A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behav Ecol Sociobiol*. 2011;65:13-21. doi:10.1007/s00265-010-1037-6

36. Bilinski A, Salomon JA, Giardina J, Ciaranello A, Fitzpatrick MC. Passing the test: a model-based analysis of safe school-reopening strategies. *Ann Intern Med*. 2021;174(8):1090-1100. doi:10.7326/M21-0600

37. Bilinski A, Hatfield LA. Nothing to see here? non-inferiority approaches to parallel trends and other model assumptions. *arXiv*. Preprint posted May 8, 2018. doi:10.48550/arXiv.1805.03273

38. Whaley CM, Cantor J, Pera M, Jena AB. Assessing the association between social gatherings and COVID-19 risk using birthdays. *JAMA Intern Med*. 2021;181(8):1090-1099. doi:10.1001/jamainternmed.2021.2915

39. Suñer C, Coma E, Ouchi D, et al. Association between two mass-gathering outdoor events and incidence of SARS-CoV-2 infections during the fifth wave of COVID-19 in north-east Spain: a population-based control-matched analysis. *Lancet Reg Health Eur*. 2022;15:100337. doi:10.1016/j.lanepe.2022.100337

40. Szablewski CM, Chang KT, Brown MM, et al. SARS-CoV-2 transmission and infection among attendees of an overnight camp—Georgia, June 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(31):1023-1025. doi:10.15585/mmwr.mm6931e1

41. Lopez AS, Hill M, Antezano J, et al. Transmission dynamics of COVID-19 outbreaks associated with child care facilities—Salt Lake City, Utah, April–July 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(37):1319-1323. doi:10.15585/mmwr.mm6937e3

42. Lam-Hine T, McCurdy SA, Santora L, et al. Outbreak associated with SARS-CoV-2 B.1.617.2 (Delta) variant in an elementary school—Marin County, California, May-June 2021. *MMWR Morb Mortal Wkly Rep*. 2021;70(35):1214-1219. doi:10.15585/mmwr.mm7035e2

43. Pray IW, Gibbons-Burgener SN, Rosenberg AZ, et al. COVID-19 outbreak at an overnight summer school retreat—Wisconsin, July–August 2020. *MMWR Morb Mortal Wkly Rep*. 2020;69(43):1600-1604. doi:10.15585/mmwr.mm6943a4

44. Doyle T, Kendrick K, Troelstrup T, et al. COVID-19 in primary and secondary school settings during the first semester of school reopening—Florida, August–December 2020. *MMWR Morb Mortal Wkly Rep*. 2021;70(12):437-441. doi:10.15585/mmwr.mm7012e2

45. Vlachos J, Hertegård E, B Svaleryd H. The effects of school closures on SARS-CoV-2 among parents and teachers. *Proc Natl Acad Sci U S A*. 2021;118(9):e2020834118. doi:10.1073/pnas.2020834118

46. Head JR, Andrejko KL, Cheng Q, et al. School closures reduced social mixing of children during COVID-19 with implications for transmission risk and school reopening policies. *J R Soc Interface*. 2021;18(177):20200970. doi:10.1098/rsif.2020.0970

47. Ludvigsson JF, Engerström L, Nordenhäll C, Larsson E. Open schools, COVID-19, and child and teacher morbidity in Sweden. *N Engl J Med*. 2021;384(7):669-671. doi:10.1056/NEJMc2026670

48. Ludvigsson JF. Systematic review of COVID-19 in children shows milder cases and a better prognosis than adults. *Acta Paediatr*. 2020;109(6):1088-1095. doi:10.1111/apa.15270

49. Leidman E, Duca LM, Omura JD, Proia K, Stephens JW, Sauber-Schatz EK. COVID-19 trends among persons aged 0–24 years—United States, March 1–December 12, 2020. *MMWR Morb Mortal Wkly Rep*. 2021;70(3):88-94. doi:10.15585/mmwr.mm7003e1

50. Viner RM, Russell SJ, Croker H, et al. School closure and management practices during coronavirus outbreaks including COVID-19: a rapid systematic review. *Lancet Child Adolesc Health*. 2020;4(5):397-404. doi:10.1016/S2352-4642(20)30095-X

51. COVID-19 Vaccines. U.S. Department of Health and Human Services. 2022. Accessed April 5, 2023. https://www.hhs.gov/coronavirus/covid-19-vaccines/index.html

52. Karlamangla S. These schools will require masks even after California's mandate ends. *New York Times*. March 10, 2022. Accessed February 1, 2023. https://www.nytimes.com/2022/03/10/us/california-masks-schools.html

53. Irfan O, Li J, Tang K, Wang Z, Bhutta ZA. Risk of infection and transmission of SARS-CoV-2 among children and adolescents in households, communities and educational settings: a systematic review and meta-analysis. *J Glob Health*. 2021;11:05013. doi:10.7189/jogh.11.05013

54. Fisman DN, Greer AL, Brankston G, et al. COVID-19 case age distribution: correction for differential testing by age. *Ann Intern Med*. 2021;174(10):1430-1438. doi:10.7326/M20-7003

55. Will M. California mandates that teachers get vaccinated or regularly tested for COVID. EducationWeek. August 11, 2021. Updated August 12, 2021. Accessed April 5, 2023. https://www.edweek.org/teaching-learning/california-mandates-that-teachers-get-vaccinated-or-regularly-tested-for-covid/2021/08

SUPPLEMENT 1.

eAppendix.

eFigure 1. Directed Acyclic Graphs (DAGs) Representing Instrumental Variable Analyses With Various Types of Measurement Error
eFigure 2. SEIR Model Used to Simulate Data According to the Hypothesized Data Generating Mechanism
eTable 1. Parameter Values Used in the Simulation of Data
eTable 2. Scenarios From eFigure 1 Were Examined Using the Following Combinations of Parameters
eFigure 3. Comparison of Observed and True Incidence Rate Ratios (A) and True and Adjusted Incidence Rate Rates (B)
eFigure 4. Heat Map Comparison of Observed and True Incidence Rate Ratios (A) and True and Adjusted Incidence Rate Rates (B)
eFigure 5. Timeline of Circulating SARS-CoV-2 Variants, Vaccination Approval Dates, and School Periods
eFigure 6. Distribution of Testing Rate Ratios and Test Positivity Rate Ratios Among California Counties Comparing Testing Rates and Test Positivity Rates Among School-Aged Children (5-10 Years) to Non-School-Aged Children (0-4 Years)
eFigure 7. COVID-19 Incidence as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models Assuming a Quadratic Relationship Between Age and Incidence
eFigure 8. COVID-19 Incidence as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models That Use Local Linear Regression
eFigure 9. County-Specific (Black) and Pooled (Blue) Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 in the Fall 2021 Semester of the 2021-2022 Academic Year Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After
eFigure 10. County-Specific (Black) and Pooled (Blue) Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 in the Spring 2022 Semester of the 2021-2022 Academic Year Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After
eFigure 11. County-Specific (Black) and Pooled (Blue) Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 in the Fall 2022 Semester of the 2021-2022 Academic Year Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After
eFigure 12. Comparison of IRRs During In-School Periods Adjusting for Testing Biases and Not Adjusting for Testing Differences
eFigure 13. Pooled Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After
eFigure 14. COVID-19 Hospitalization as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models Assuming a Linear Relationship Between Age and Hospitalizations
eFigure 15. COVID-19 Hospitalization as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models Assuming a Quadratic Relationship Between Age and Hospitalizations
eFigure 16. COVID-19 Hospitalization as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models That Use Local Linear Regression
eFigure 17. Associations Between Elementary School Age-Eligibility and Hospitalizations for COVID-19 by School Period and Model Parameterization
eFigure 18. Power Analysis for the Association Between Hospitalizations and School Eligibility
eTable 3. Total Number of Cases and Hospitalizations Among the Subsample of Children Who Fell Within 24 Months, in Either Direction, of the Elementary School Attendance Threshold
eTable 4. Incidence Rate Ratios (IRRs) and 95% Confidence Intervals Comparing the Incidence of COVID-19 Among Children Born Just Before the Threshold for Elementary School Attendance (September 1st) Compared to Just After
eTable 5. Results of Meta-Analysis
eReferences.

SUPPLEMENT 2.
Data Sharing Statement

Supplementary Online Content

Lin E, Bilinski A, Collender PA, et al. COVID-19 incidence and age eligibility for elementary school.

JAMA Netw Open. 2024;7(11):e2444836. doi:10.1001/jamanetworkopen.2024.44836

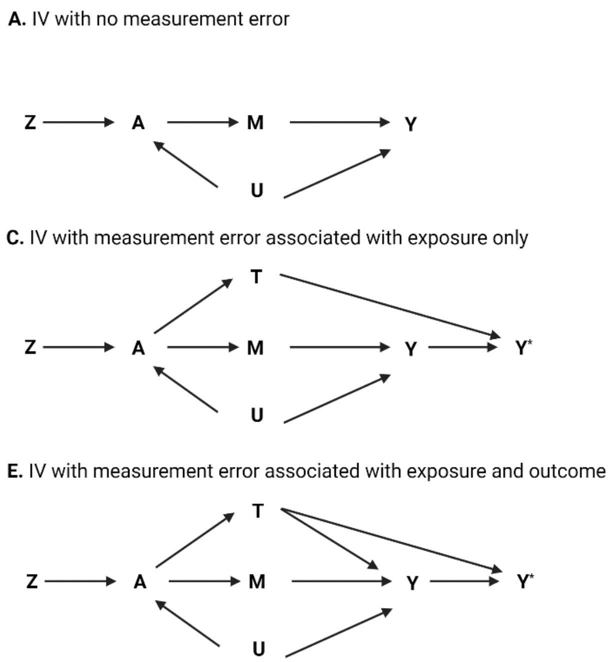
This supplementary material has been provided by the authors to give readers additional information about their work.

eAppendix.

Simulation study- Methods

This study uses reported COVID-19 cases, which are imperfectly ascertained by reportable disease surveillance systems. We conducted a simulation study in order to understand how measurement error of the outcome biases the estimated incidence rate ratios, and to estimate an adjustment factor that reduces this bias.

First, we drew directed acyclic graphs (DAGs) to visually depict the possible ways in imperfect outcome measurement (here, due to lack of testing) might occur within the study (eFigure 1). These scenarios included: no measurement error; measurement error unassociated with the exposure and the true outcome; measurement error associated with the outcome only or the exposure only; and measurement error associated with both the exposure and the outcome. In this particular case of COVID-19, we believe that measurement error could be associated with both the exposure (higher testing in school aged populations) and the true outcome (case ascertainment leads to case isolation and reduction in cases).

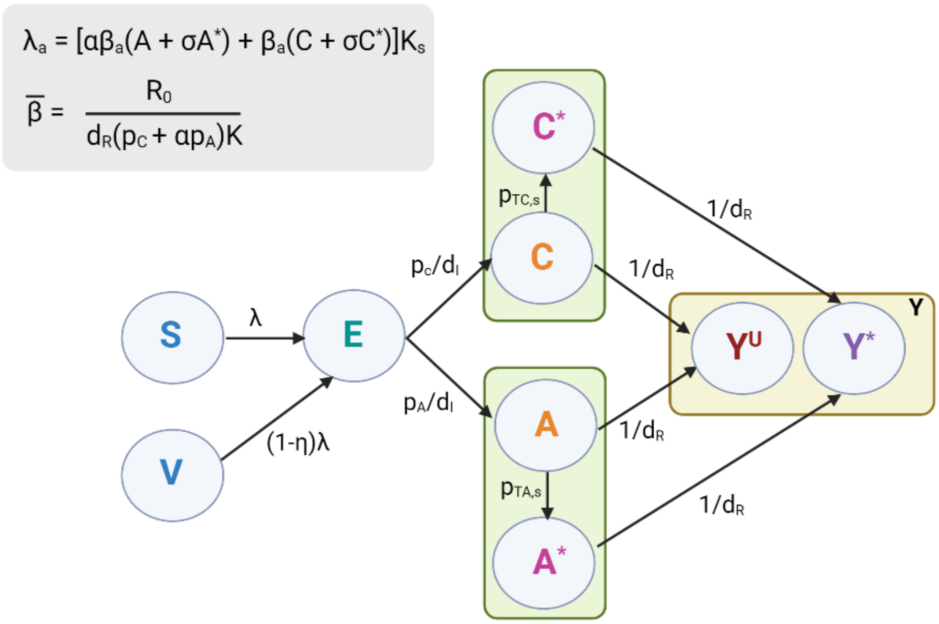


eFigure 1. Directed Acyclic Graphs (DAGs) Representing Instrumental Variable Analyses With Various Types of Measurement Error. IV = instrumental variable; Z = running variable (child's birthdate); A = exposure (school attendance); M = mediators (e.g., child social networks inside and outside of classrooms, adult social networks, vaccination willingness, etc.); U = unmeasured confounders (e.g., community controls); Y = true outcome (all COVID cases); Y* = measured outcome (reported COVID cases); T = Testing/case detection.

Next, we simulated data using a compartmental SEIR model (eFigure 2) for each data generating mechanism depicted in eFigure 1 and using the parameter values shown in eTable 1. The SEIR model included compartments for asymptomatic versus symptomatic cases, and tested (i.e., reported) cases versus not tested cases (i.e., not observed). Movement from the susceptible to the exposed compartment was proportional to the force of infection, which was defined as being conditional on age (a; via pathogen transmission rate (β) and probability case is clinical ($p_{C,a}$)) and school attendance (S; via contact rates, K_s, and case detection, $p_{TC,s}$ and $p_{TA,s}$) according to the following equations:

$$\lambda_{a,s} = [\alpha\beta_a(A + \sigma A^*) + \beta_a(C + \sigma C^*)]K_S \qquad [1]$$

$$\beta_a = \frac{R_0}{d_R[\alpha p_{A,a}+P_{C,a}]K_S} + 0.0001a \qquad [2]$$



eFigure 2. SEIR Model Used to Simulate Data According to the Hypothesized Data Generating Mechanism. S = susceptible; V = vaccinated; E = exposed; C = clinical case; C* = reported clinical case; A = asymptomatic case; A* = reported asymptomatic case; Y = total recovered; Y^U = unobserved total recovered; Y^*= observed total recovered. Because we simulate data over a short period of time (one semester), for this example, we assume some children start the semester vaccinated, and that no waning immunity will occur.

eTable 1. Parameter Values Used in the Simulation of Data

Parameter	Definition	Value
a	Age, in months, centered at 0	Varied in simulation from -24 to 24
$\lambda_{a,s}$	School, and age-dependent force of infection	Calculated (equation 1)
β_a	Age-dependent transmission rate of pathogen	Calculated (equation 2)
R_0	Basic reproduction number	2.5
η	Vaccine effectiveness	0.5
d_I	Incubation period	2 days
d_R	Infectious period	5 days
$p_{C,a}$	Age-dependent probability case is clinical	0.3 + 0.0001*a
$p_{A,a}$	Age-dependent probability case is asymptomatic	$1 - p_C$
α	Relative infectivity of asymptomatic case to symptomatic case	0.5
$p_{TC,s}$	School-dependent case ascertainment of clinical cases	$p_{TC,s=0} = 0.2$ $p_{TC,s=1}$ varied in simulation
$p_{TA,s}$	School-dependent case ascertainment of asymptomatic cases	$p_{TA,s=0} = 0.05$ $p_{TA,s=1}$ varied in simulation
K_S	School-dependent contact rates	$K_{S=0} = 10$ $K_{S=1}$ varied in simulation
σ	Relative contribution of tested case to transmission	Varied between 0 (full isolation) and 1 (no isolation)

Because we assumed that the primary contribution of schools to COVID-19 incidence is via $K_s, p_{TC,s}$, and $p_{TA,s}$, we first run the model for various combinations of these values at all age levels considered (here, we center age at 0 and examine within 24 months of this threshold). Specifically, we set the values equal to some baseline value for no school attendance ($S = 0$) and vary the degree that school attendance increases them. In cases where measurement is associated with the true outcome (i.e., tested cases are isolated), we assumed that reported cases contributed to the force of infection (via a parameter $\sigma < 1$) (eTable 2).

eTable 2. Scenarios From eFigure 1 Were Examined Using the Following Combinations of Parameters:

Scenario (eFigure 1)	**Case detection**	**Testing ratio** *school detection to out of school detection*	**Removal of detected cases**
A	$p_{TC,s} = 1;\ p_{TA,s} = 1$	Test ratio = 1	No removal ($\sigma = 1$)
B	$p_{TC,s} < 1;\ p_{TA,s} < 1$	Test ratio = 1	$\sigma = 1$
C	$p_{TC,s} < 1;\ p_{TA,s} < 1$	Test ratio > 1	$\sigma = 1$
D	$p_{TC,s} < 1;\ p_{TA,s} < 1$	Test ratio = 1	$\sigma < 1$
E	$p_{TC,s} < 1;\ p_{TA,s} < 1$	Test ratio > 1	$\sigma < 1$

We simulated observed cases in each age group, represented by Y_a^*, as well as true cases, represented by $Y_a = Y_a^* + Y_a^U$. We then fit the linear RD model to the simulated data using observed cases (Y_a^*) as the outcome, and true cases as the outcome (Y_a) to generate the true IRR and the observed IRR. Finally, we examined various functions of the testing ratio we should upweight the cases who did not attend school by such that an adjusted IRR approximates the true IRR.

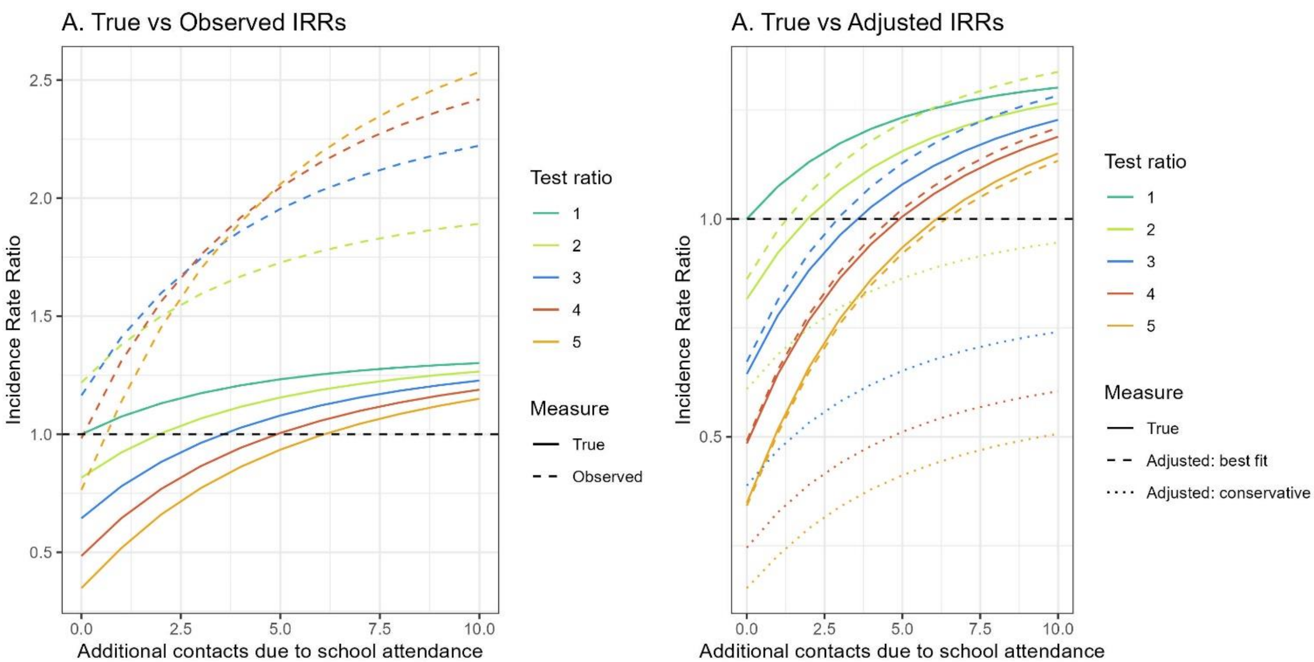
Simulation study- Results

The true IRR was equivalent to the observed IRR over all variations of school-dependent contact rate and testing ratios in the scenario where case acquisition is not associated with the exposure (school attendance; DAGs A, B, and D in eFigure 1). Of note, the true and observed absolute difference (incidence rate difference as opposed to incidence rate ratio) would be differ in the presence of measurement error (D and D in eFigure 1).

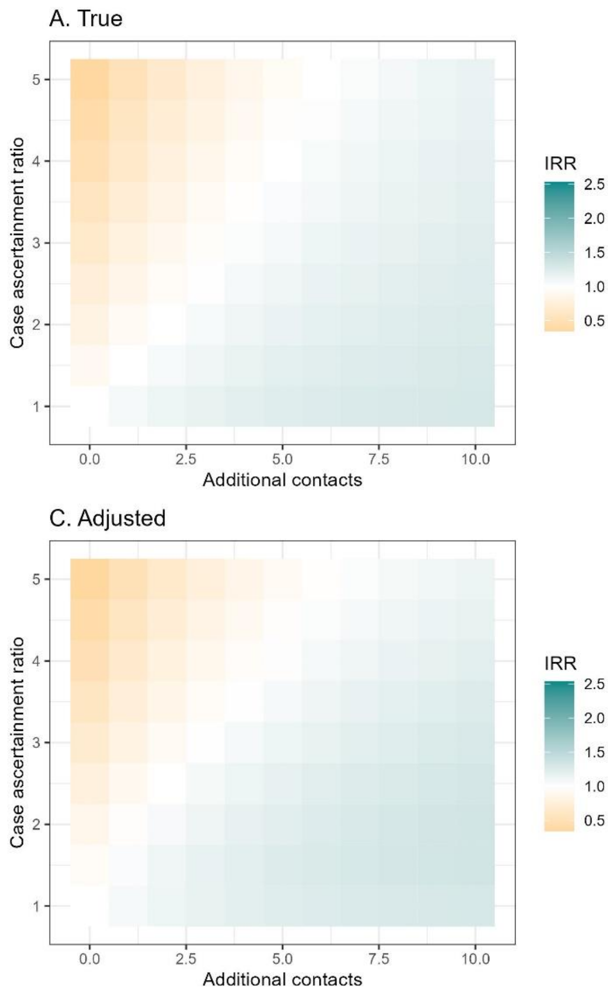
Under the scenario where case detection is associated with the exposure but not associated with the true outcome (DAG C in eFigure 1), observed IRR exceeds true IRR. This bias is intuitive: higher case detections lead greater observed IRRs. In this context, the true IRR varies when school-dependent contact rates are varied, but not when the testing rate ratio is varied.

Under the scenario where case detection is associated with the exposure as well as with the true outcome (DAG E in eFigure 1), the observed IRR exceeds true IRR. However, the bias is more complex. At lower additional school-based contacts (K), the observed IRR at a high test ratio may be lower than the observed IRR at a low test ratio due to removal of infected individuals. At higher contact rates, the relationship flips (eFigure 3A, eFigure 4A-B). Of note, this complex relationship is only observed when case detection strongly prevents cases from contributing to the force of infection ($\sigma < 0.5$)

In both scenarios where the true IRR was not equal to the observed IRR, upweighting the number of cases in the population who did not attend school by the square root of the testing ratio resulted in the best approximation of the adjusted IRR to the true IRR (eFigure 3A, eFigure 4A&C. Therefore, we used the square root of the testing rate ratio as our primary weighting factor in the main analysis.



eFigure 3. Comparison of Observed and True Incidence Rate Ratios (A) and True and Adjusted Incidence Rate Rates (B). In this example, data is simulated according to the data generating mechanism depicted in eFigure 1E. In this mechanism, testing is positively associated with the exposure and reported outcome, but negative associated with the true outcome.



eFigure 4. Heat Map Comparison of Observed and True Incidence Rate Ratios (A) and True and Adjusted Incidence Rate Rates (B). In this example, data is simulated according to the data generating mechanism depicted in eFigure 1E. In this mechanism, testing is positively associated with the exposure and reported outcome, but negative associated with the true outcome.

Adjusting for testing bias

Testing effort was typically higher among school-aged children compared to younger children (eFigure 6). To contend with bias from differential case ascertainment, we weighted cases in the school-ineligible strata in order to approximate the number of cases that would have been observed in the school-ineligible age groups given equal testing effort (Supplemental Methods). To estimate these weights, we obtained aggregate data on the number of total and positive tests by county and age group (0- to 4-year-olds and 5- to 10- year-olds). Using population denominators from the American Community Survey,[1] we calculated the number of tests per age group and calculated the school-associated testing ratio as the testing rate among the school-eligible 5- to 10- year-olds versus 0- to 4- year-olds. These values ranged from 0.54 to 4.72, with a mean of 1.33 (eFigure 6A).

We weighted cases in the school-ineligible strata by the county's testing rate ratio in children 5- to 10- years-old vs. 0- to 4- years-old by the square root of the testing rate ratio for each county, according to our simulation analysis. In more conservative analyses, we weighted the school-ineligible strata by the testing rate ratio. This strategy is considered conservative because testing was conducted as a screening tool among asymptomatic populations, so each test administered had a lower probability of being a positive. Our simulation study further demonstrates that point.

We did not apply weights for the analyses with hospitalization as the outcome, as ascertainment of severe cases is more likely to be similar between populations.

Meta-analysis

We fit separate models for 46 of California's 58 counties, excluding 12 counties with few cases and births reported. We used a meta-analysis approach that has been commonly used to combine effect estimates across multiple locations, while permitting examination of effect heterogeneity.[2] We computed τ and the standard error on τ for each school period and county in California and used a fixed-effects meta-analysis to compute the pooled effect of τ for each school period.[2] The pooled effect was calculated as the weighted average of the individual county effects, where weights were equal to the inverse of the estimates' variances. In this way, more populous counties were generally assigned higher weights. We exponentiated the pooled estimate of τ to arrive at the pooled IRR.

To understand the effect of mitigation measures and other community-level factors on the impact of school attendance, we fit meta-regressions for each of the in-person semesters with predictors including county size, population density,[1] racial/ethnicity composition, various measures of social vulnerability,[3] cumulative COVID-19 incidence at the start of the school semester (as a proxy for natural immunity),[4] vaccination coverage at the start of the school semester,[5] and the percent of survey respondents who reported never wearing a mask[6] (eTable 5). Predictors were tested in univariate models and in models adjusting for total county population. We also fit one meta-regression with all in-school periods, adding a fixed effects for semester and univariate predictors described.

Sensitivity Analyses

To assess the robustness of results to model specification, we ran analyses varying bandwidth (h) from 8 to 24 months and testing three different functional bases for the relationship of age (x_i) to COVID-19 outcomes and its modification by school eligibility. These included linear and quadratic models, as well as generalized additive models in which locally linear, or LOESS, curves were used to relate age to the outcome.[7] For the local linear regression, we did not include an interaction term, as the LOESS fit should be flexible enough to capture functional differences in the relationship between age and COVID-19 outcomes above and below the school age-eligibility threshold. We compared model fit using the Akaike Information Criterion (AIC), a method for model selection that allows for comparison of likelihood-based goodness-of-fit metrics, while penalizing for additional model parameters.[8,9]

Power Simulation

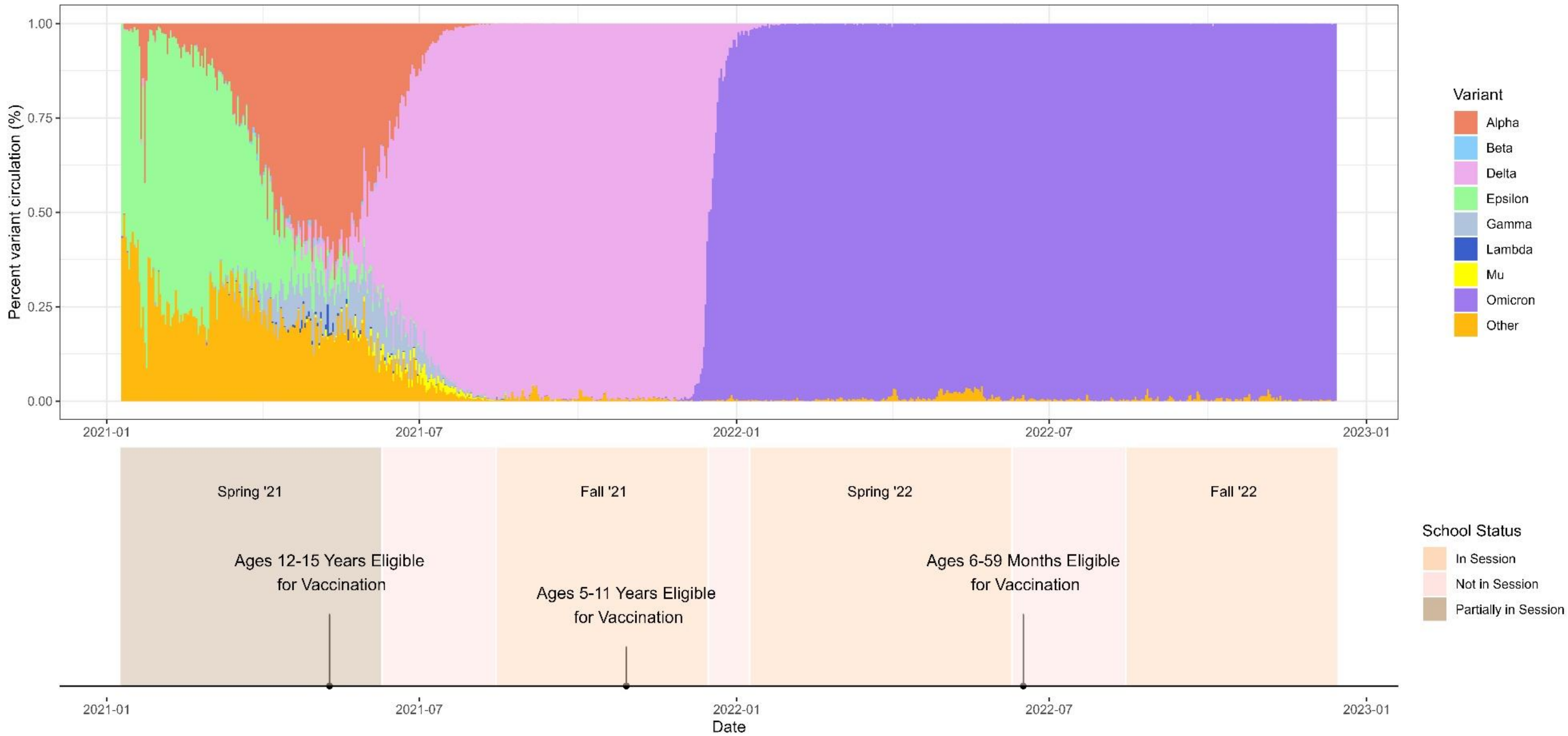
We conducted a power analysis to determine the minimum effect size for the relationship between school attendance and hospitalizations at which we would have 80% power and 95% confidence. Following the approach outlined by Bilinski and Hatfield,[10] we first fit the Poisson regression discontinuity model outlined in equation 1 (main text), using hospitalizations as the outcome and assuming a linear relationship between age and hospitalizations. We extracted model residuals, u_i, and generated predicted values, $\widehat{p}_i$, from this model, setting Z_i equal to 0 to simulate no effect of school eligibility.

We then generated 10,000 synthetic data sets using the following data-generating process. For each of the 10,000 iterations, we drew a value for $\log(\widehat{y_\iota})$ from a normal distribution centered at:

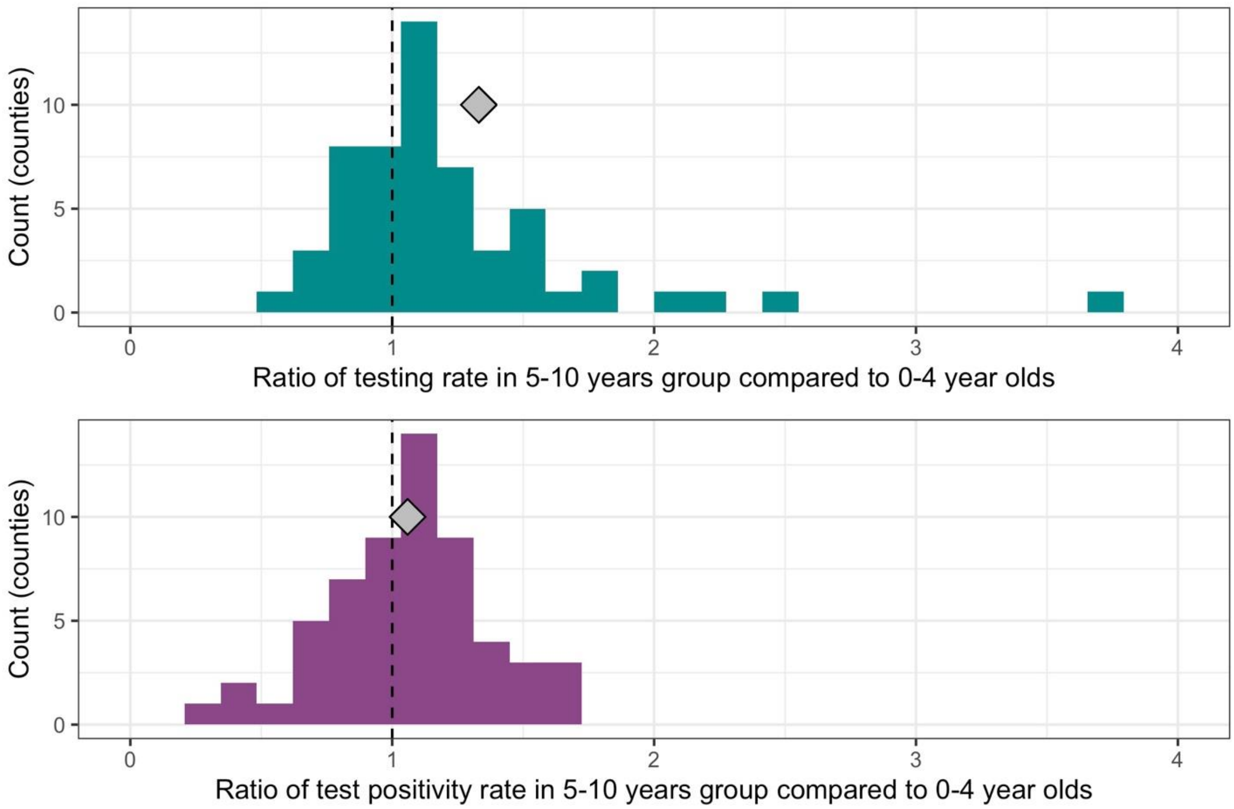
$$\log(\widehat{y_\iota}) = \widehat{p}_i + \tau Z_i$$

and with standard errors equal to the squared model residuals. We examined values for τ where $\tau \in \{0.6,1.2\}$.

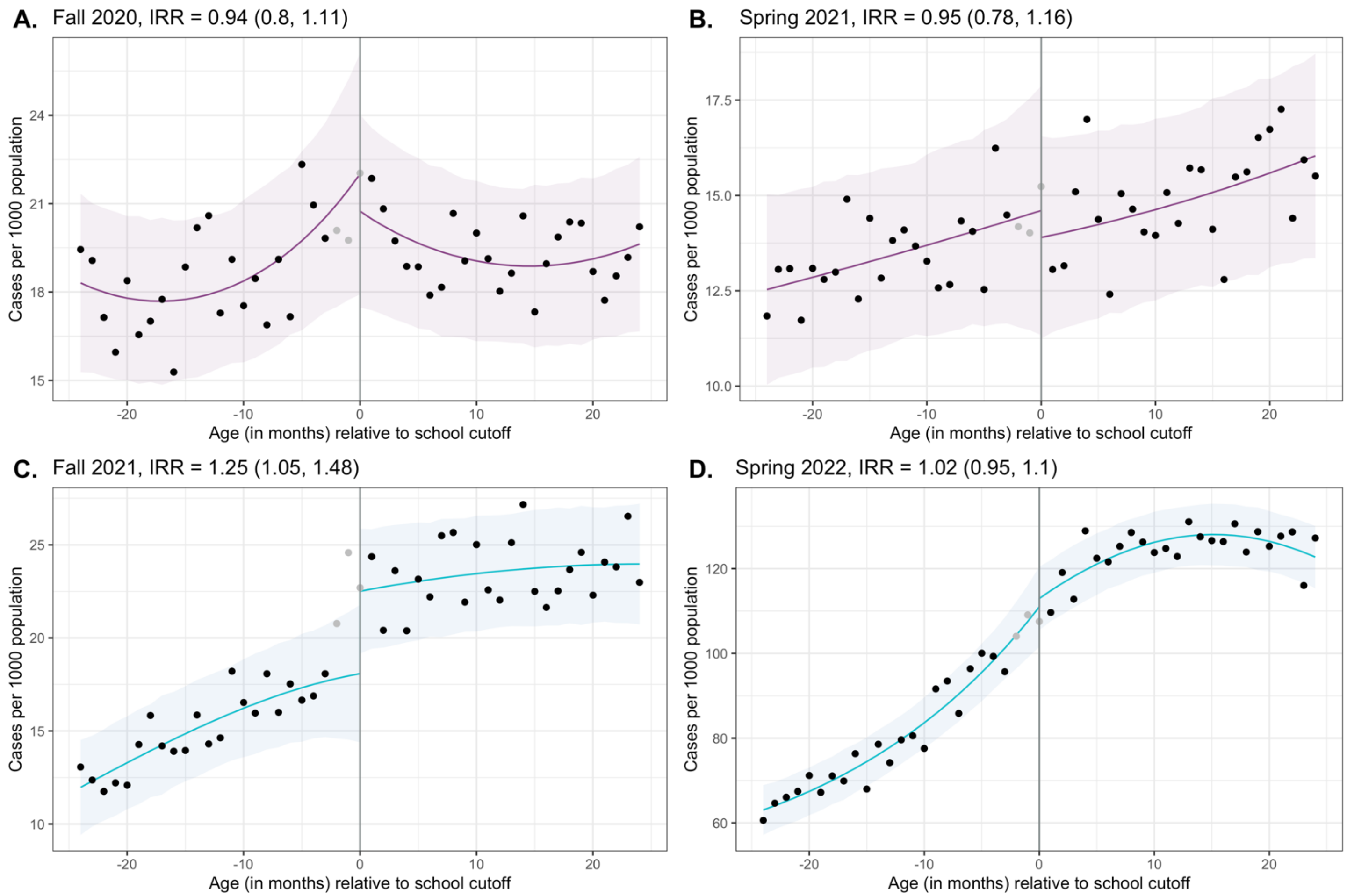
We then re-fit the Poisson regression discontinuity model to the simulated datasets, and determined whether the effect estimate on τ was statistically significant for a significance level of 0.95. Power was calculated as the proportion of the 10,000 simulations where the effect estimate was significant at the 95% confidence level. We exponentiated τ to observe the IRRs at which we had 80% power.



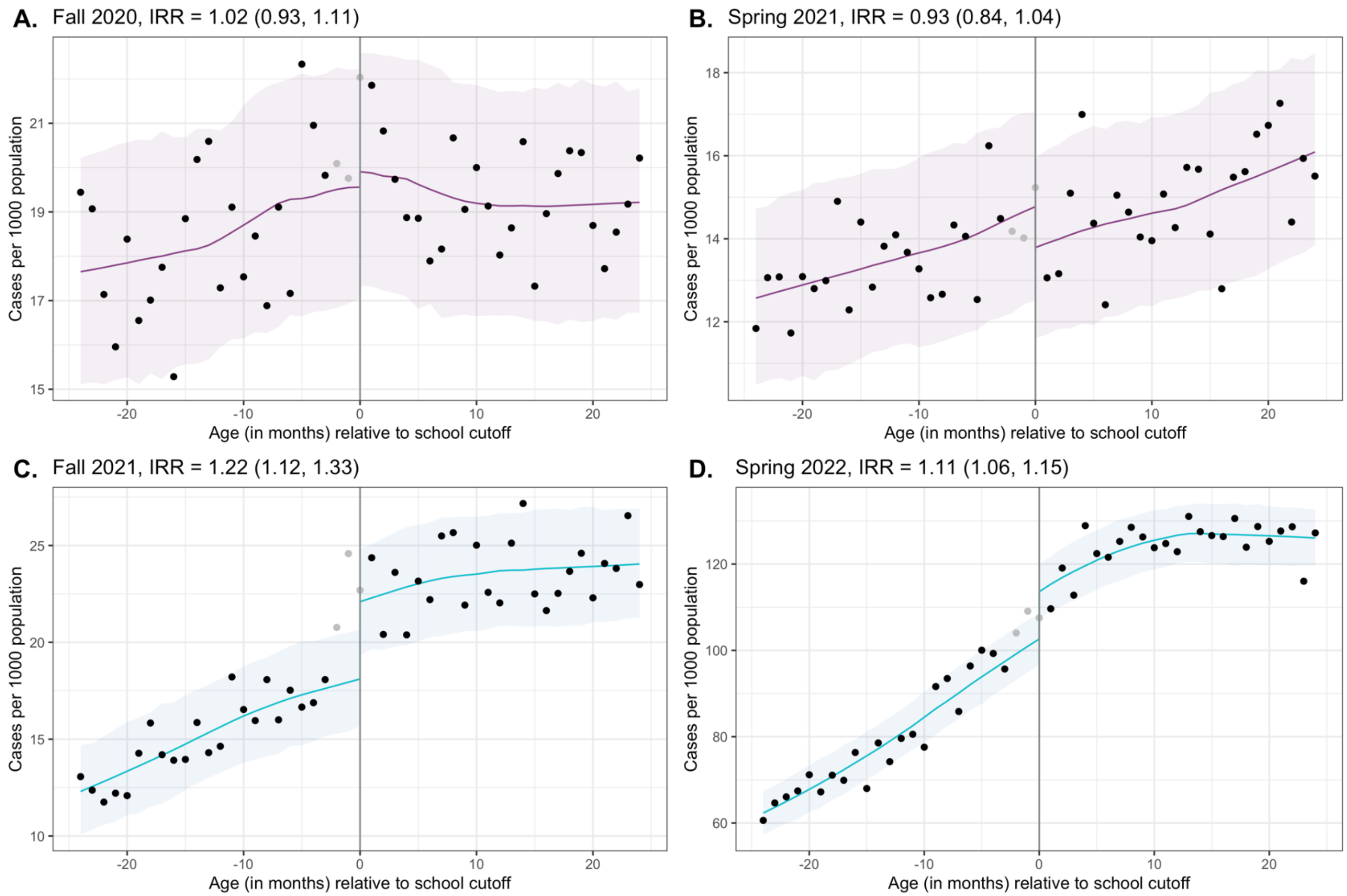
eFigure 5. Timeline of Circulating SARS-CoV-2 Variants, Vaccination Approval Dates, and School Periods. Variant data from [11]



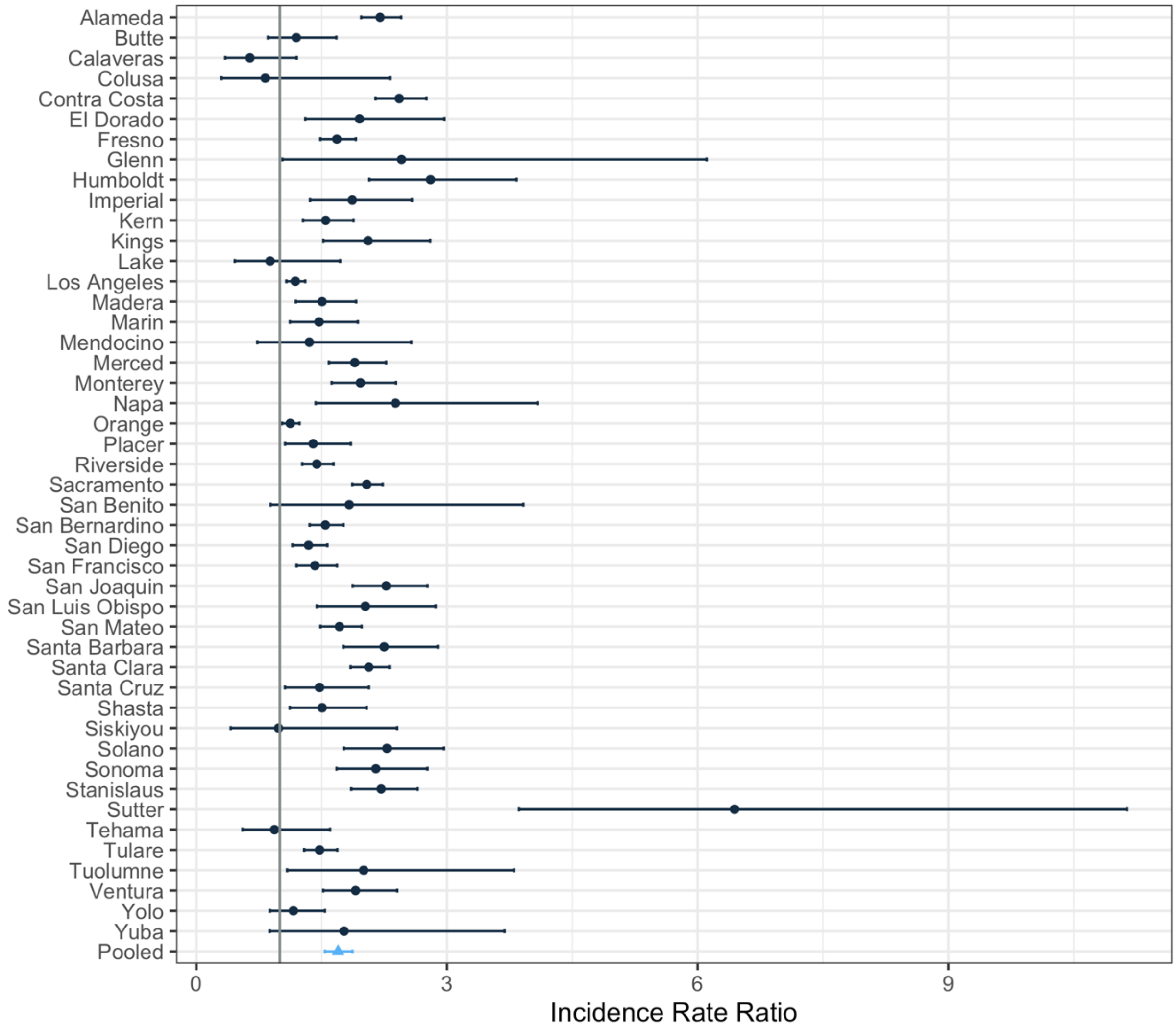
eFigure 6. Distribution of Testing Rate Ratios and Test Positivity Rate Ratios Among California Counties Comparing Testing Rates and Test Positivity Rates Among School-Aged Children (5-10 Years) to Non-School-Aged Children (0-4 Years). Vertical line indicates event testing and test positivity rate, and gray diamonds symbolize the mean ratio.



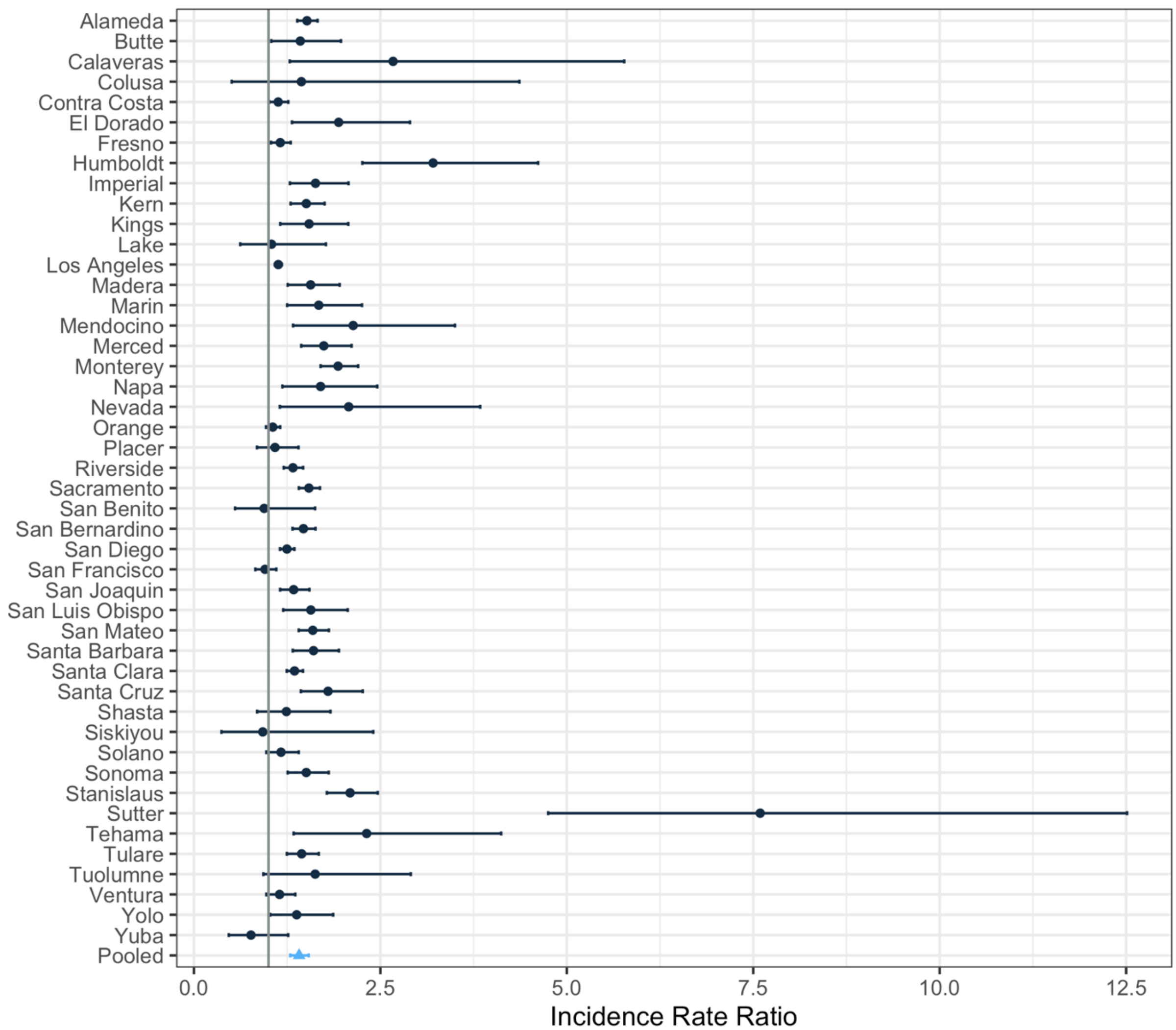
eFigure 7. COVID-19 Incidence as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models Assuming a Quadratic Relationship Between Age and Incidence. Ages right of the threshold indicate that the child is age-eligible to attend elementary school (K-5). Model fits are shown for the fall (A) and spring (B) semesters when school was remote (purple colors) and during the fall (C) and spring (D) semesters when school was in-person (blue colors). Black dots indicate observed data, lines indicate model fit, and shaded region indicates 95% prediction intervals. For this example, weighting to adjust for testing biases is not performed.Plots shown are selected from Los Angeles County, the most populous county in California, and models shown use a bandwidth of 24 months and drop children eligible for transitional kindergarten (TK).



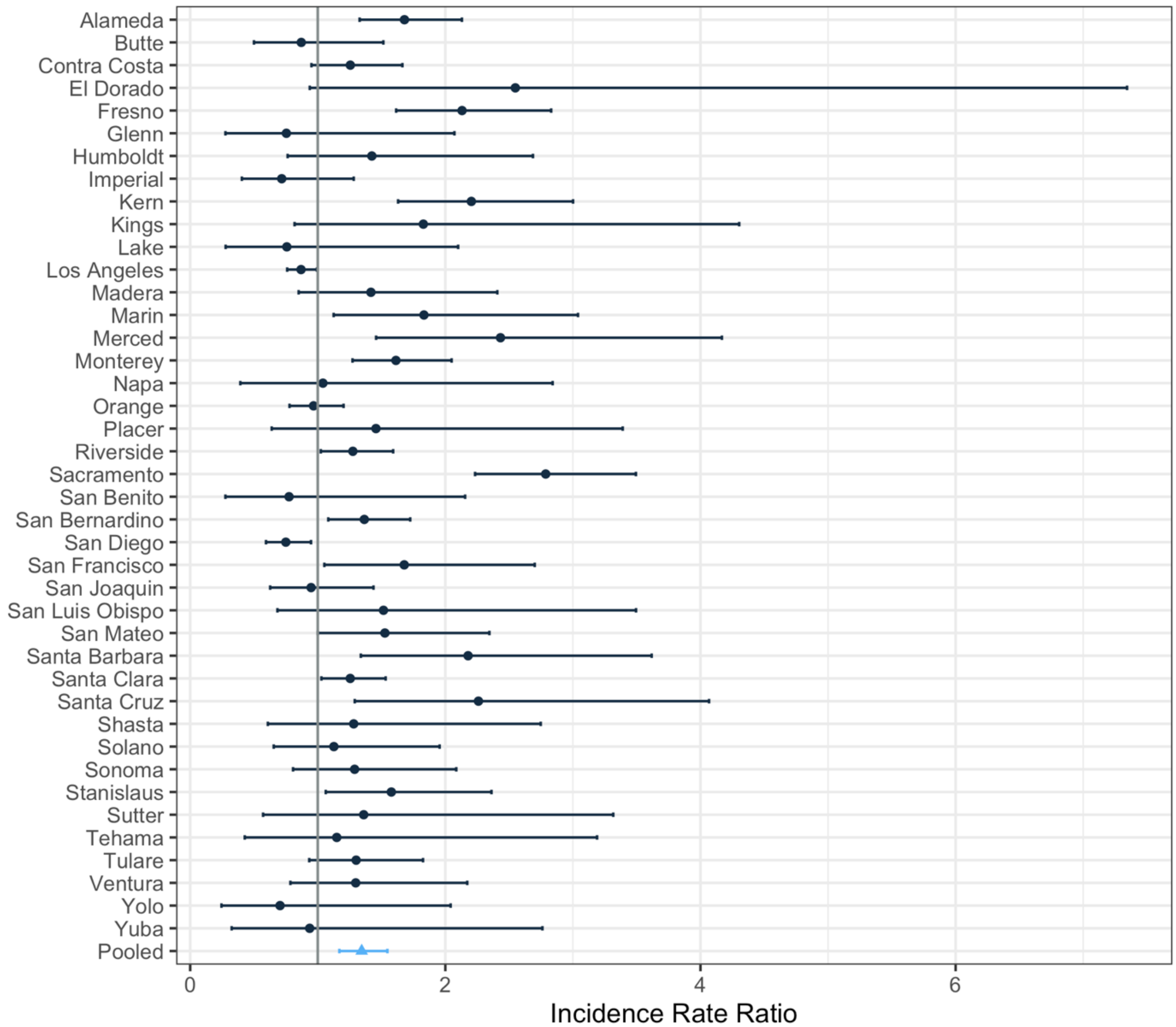
eFigure 8. COVID-19 Incidence as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models That Use Local Linear Regression. **Ages right of the threshold indicate that the child is age-eligible to attend elementary school (K-5). Model fits are shown for the fall (A) and spring (B) semesters when school was remote (purple colors) and during the fall (C) and spring (D) semesters when school was in-person (blue colors). Black dots indicate observed data, lines indicate model fit, and shaded region indicates 95% prediction intervals. For this example, weighting to adjust for testing biases is not performed. Plots shown are selected from Los Angeles County, the most populous county in California, and models shown use a bandwidth of 24 months and drop children eligible for transitional kindergarten (TK).**



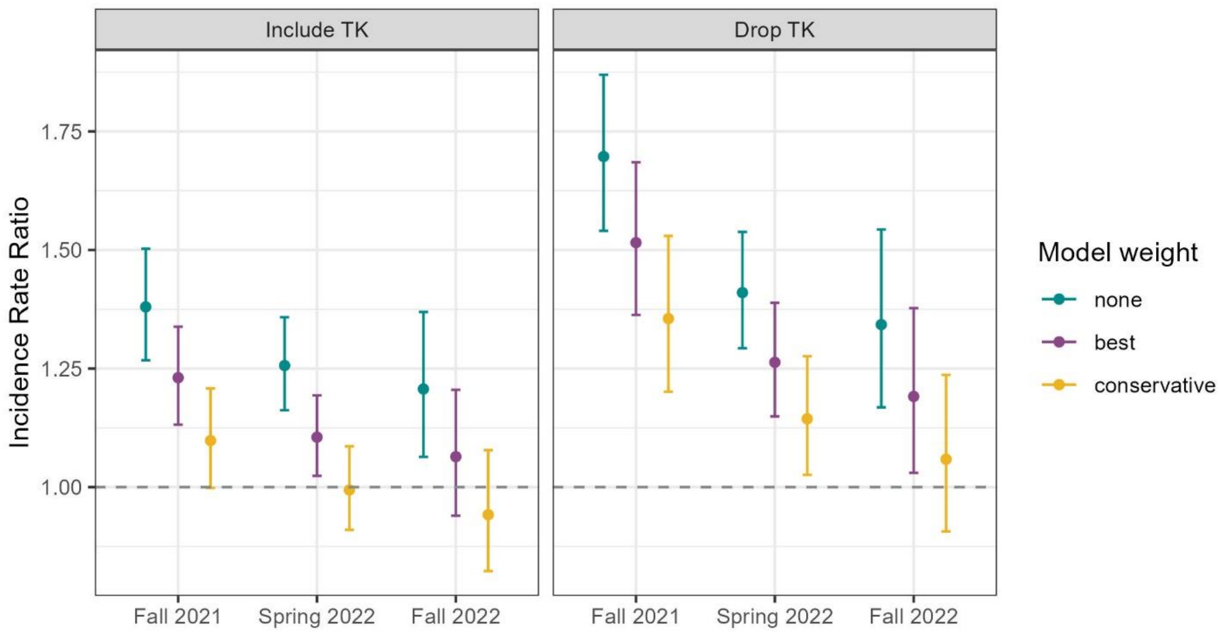
eFigure 9. County-Specific (Black) and Pooled (Blue) Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 in the Fall 2021 Semester of the 2021-2022 Academic Year Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After. **Results are for linear models with a bandwidth of 24 months, dropping children eligible for transitional kindergarten (TK).**



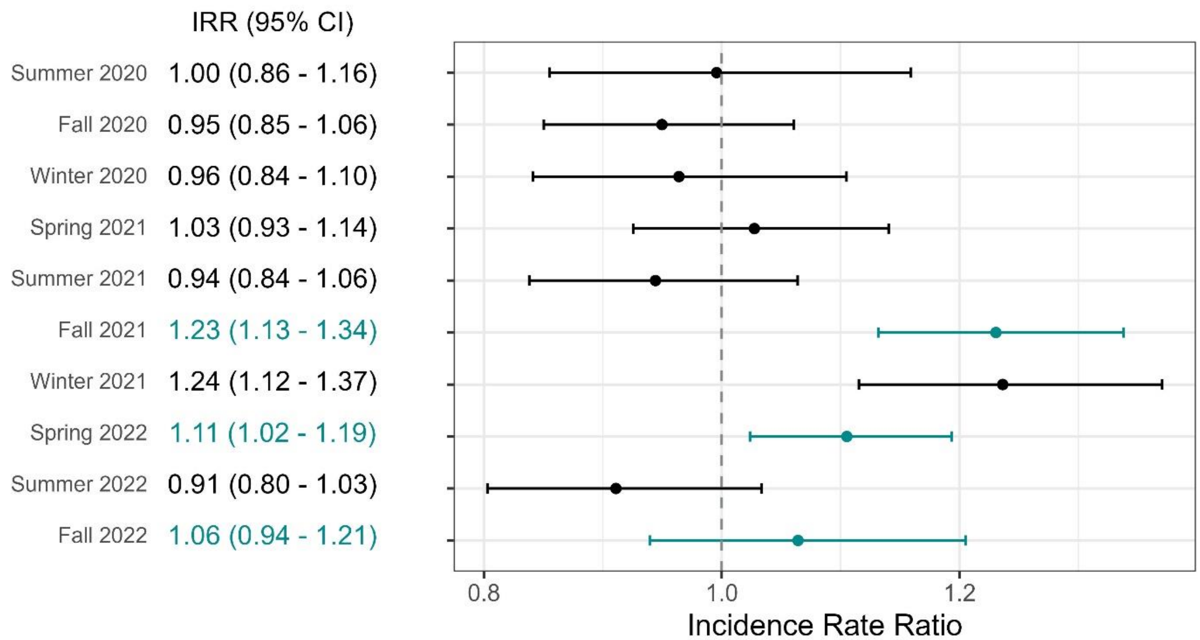
eFigure 10. County-Specific (Black) and Pooled (Blue) Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 in the Spring 2022 Semester of the 2021-2022 Academic Year Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After. **Results are for linear models with a bandwidth of 24 months, dropping children eligible for transitional kindergarten (TK).**



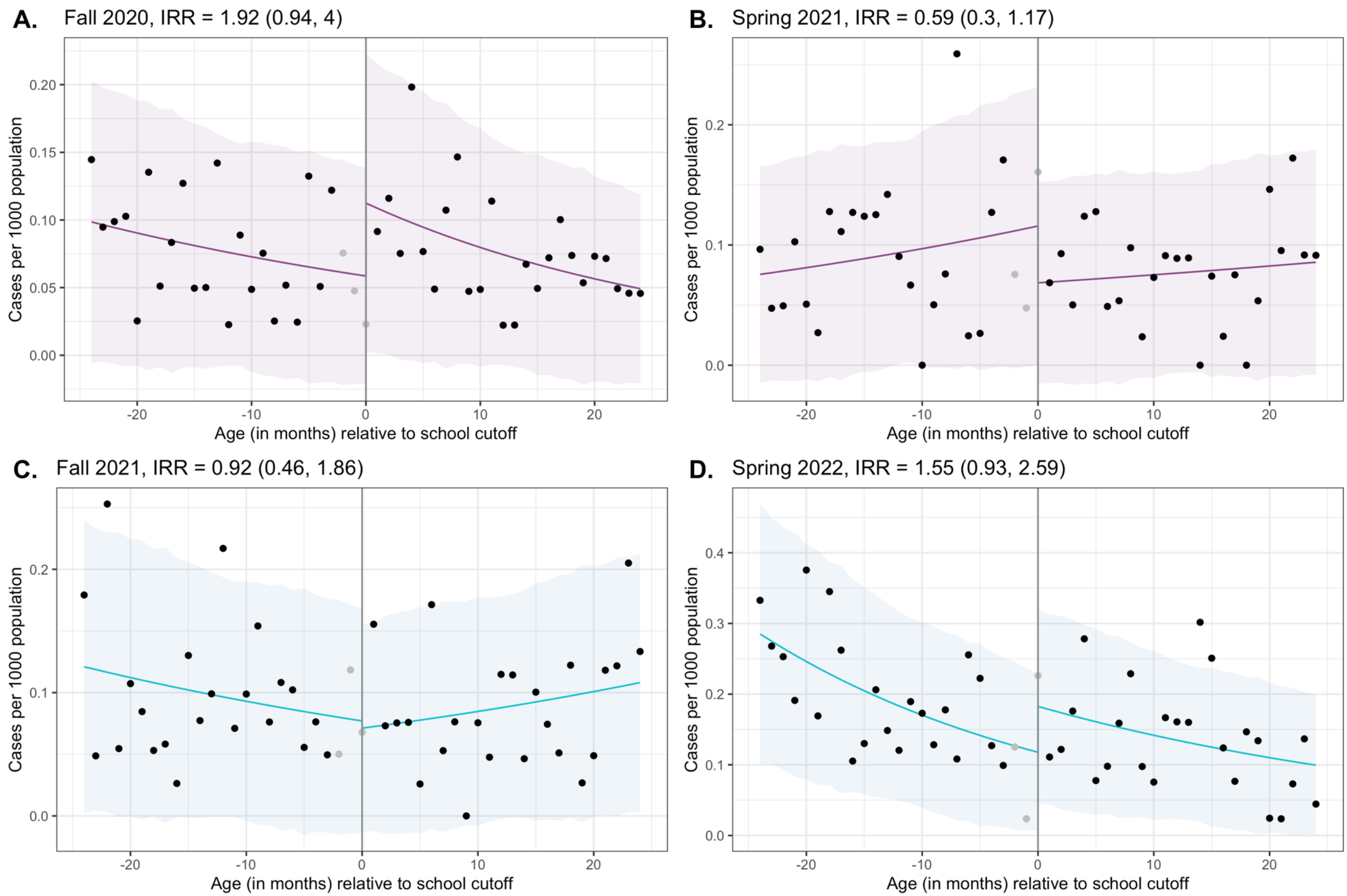
eFigure 11. County-Specific (Black) and Pooled (Blue) Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 in the Fall 2022 Semester of the 2021-2022 Academic Year Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After. **Results are for linear models with a bandwidth of 24 months, dropping children eligible for transitional kindergarten (TK).**



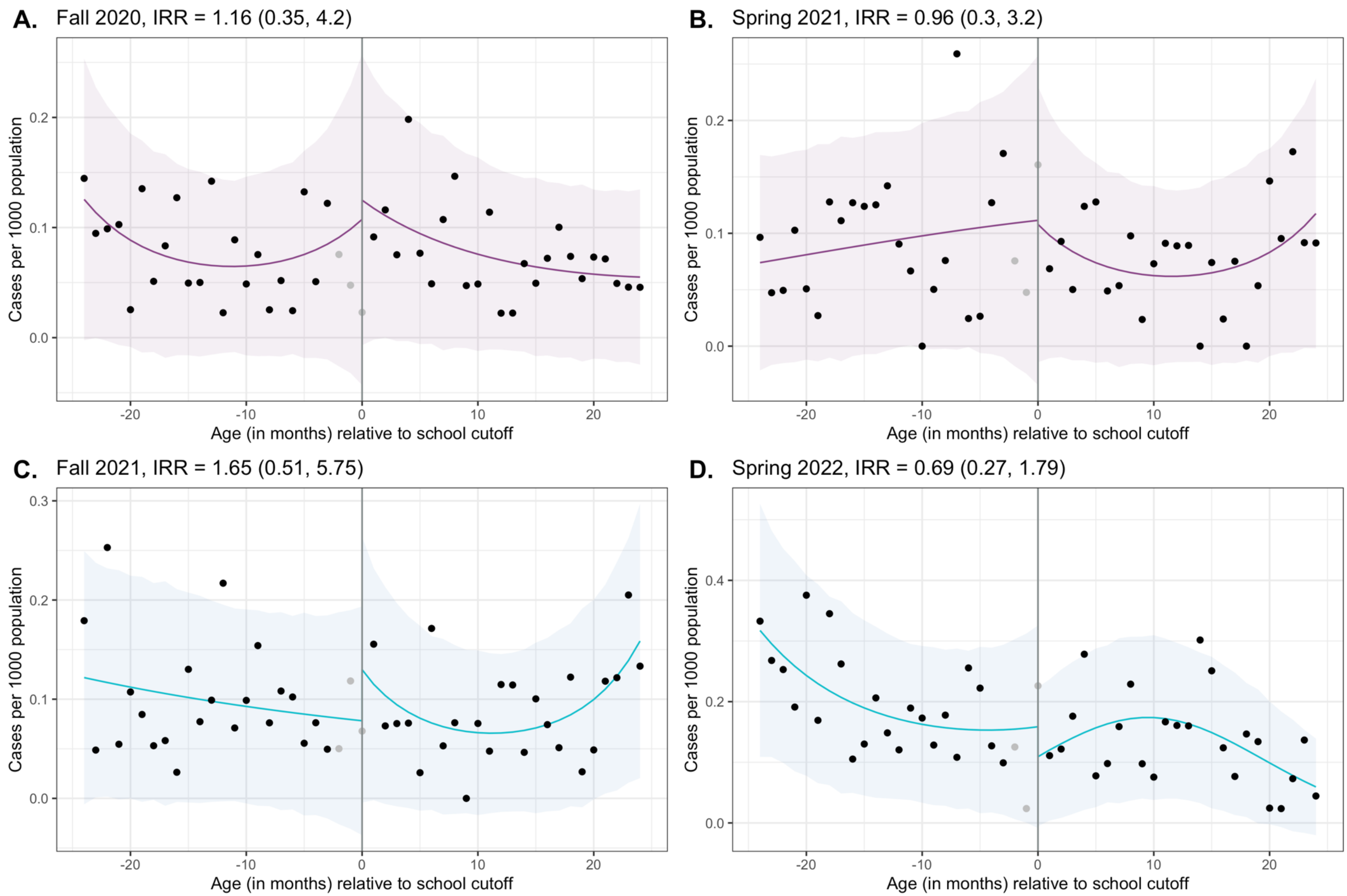
eFigure 12. Comparison of IRRs During In-School Periods Adjusting for Testing Biases and Not Adjusting for Testing Differences. Results are for linear models with a bandwidth of 24 months.



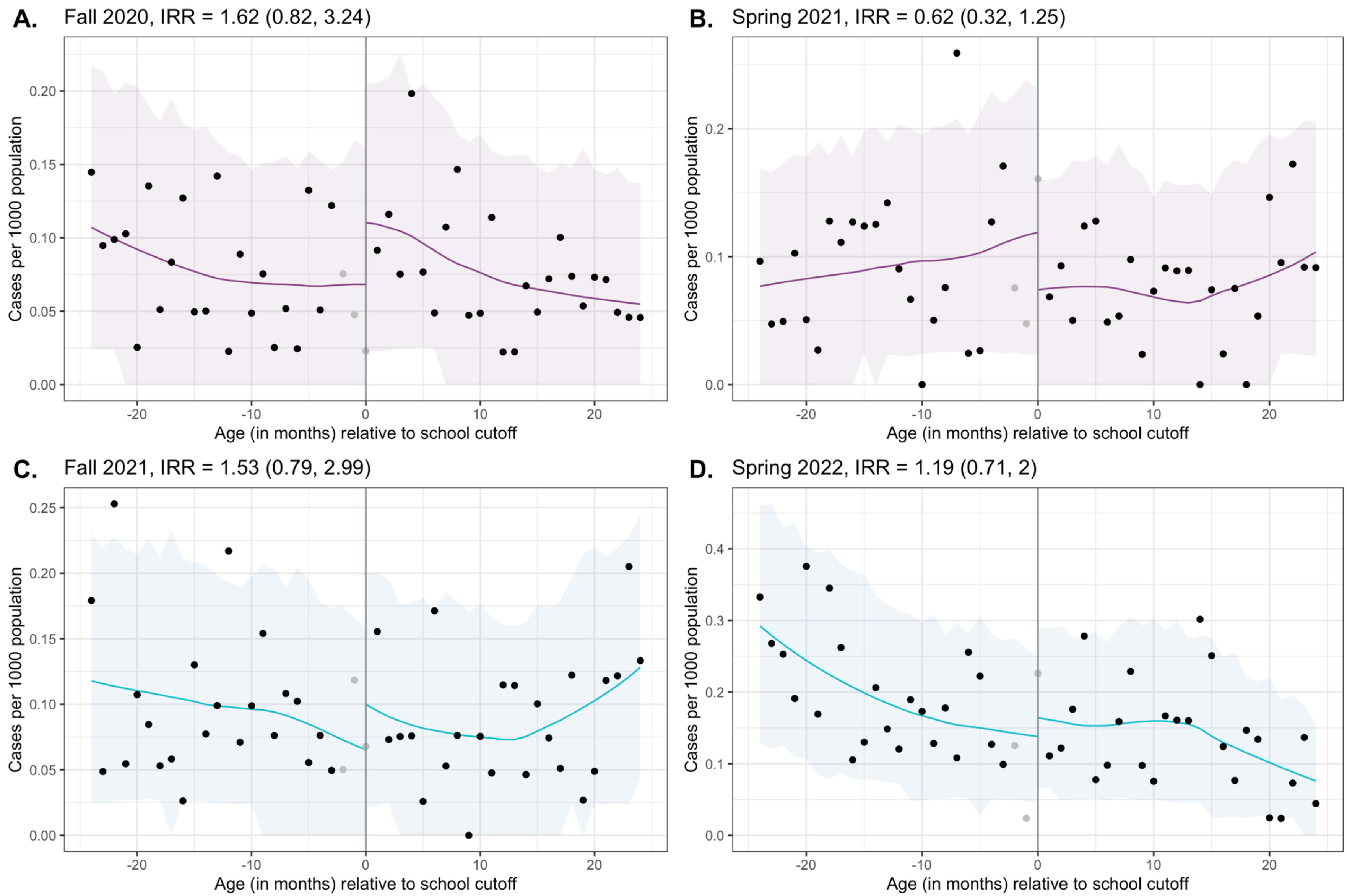
eFigure 13. Pooled Incidence Rate Ratios (IRR) Representing the Incidence of COVID-19 Among Children Born Just Before the Age-Eligibility Threshold for Elementary School Compared to Those Born Just After. Results are for linear models with a bandwidth of 24 months, including children eligible for transitional kindergarten (TK), and using the best weighting adjustment.



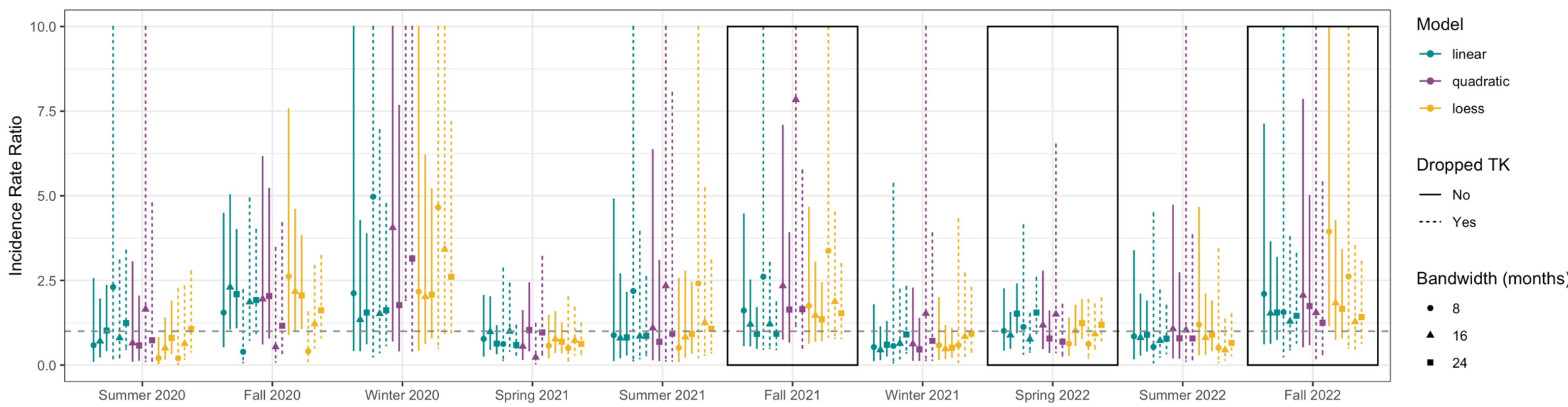
eFigure 14. COVID-19 Hospitalization as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models Assuming a Linear Relationship Between Age and Hospitalizations. Ages right of the threshold indicate that the child is age-eligible to attend elementary school (K-5). Model fits are shown for the fall (A) and spring (B) semesters when school was remote (purple colors) and during the fall (C) and spring (D) semesters when school was in-person (blue colors). Black dots indicate observed data, lines indicate model fit, and shaded region indicates 95% prediction intervals. Models shown use a bandwidth of 24 months and drop children eligible for transitional kindergarten (TK).



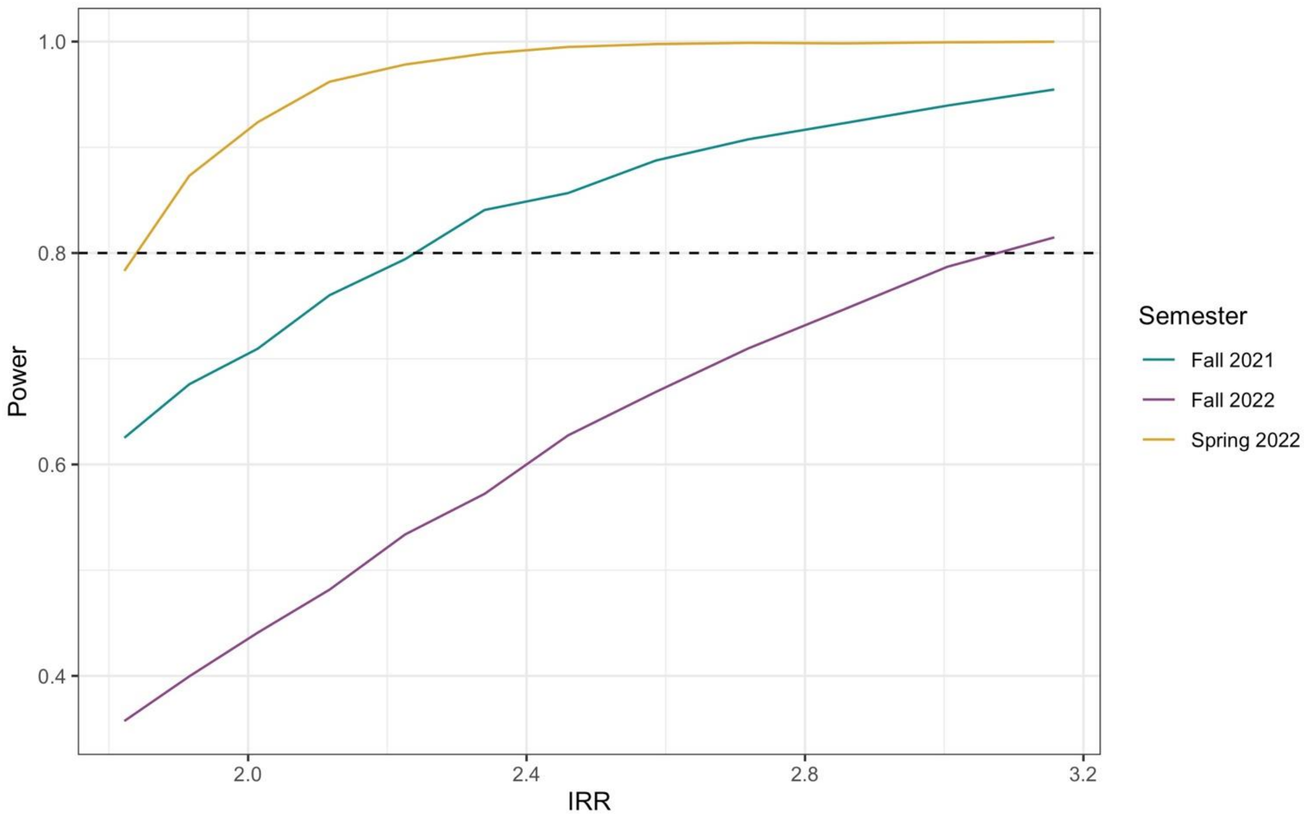
eFigure 15. COVID-19 Hospitalization as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models Assuming a Quadratic Relationship Between Age and Hospitalizations. Ages right of the threshold indicate that the child is age-eligible to attend elementary school (K-5). Model fits are shown for the fall (A) and spring (B) semesters when school was remote (purple colors) and during the fall (C) and spring (D) semesters when school was in-person (blue colors). Black dots indicate observed data, lines indicate model fit, and shaded region indicates 95% prediction intervals. Models shown use a bandwidth of 24 months and drop children eligible for transitional kindergarten (TK).



eFigure 16. COVID-19 Hospitalization as a Function of Child Age Relative to the September 1st Cutoff for Elementary School Attendance, for Models That Use Local Linear Regression. Ages right of the threshold indicate that the child is age-eligible to attend elementary school (K-5). Model fits are shown for the fall (A) and spring (B) semesters when school was remote (purple colors) and during the fall (C) and spring (D) semesters when school was in-person (blue colors). Black dots indicate observed data, lines indicate model fit, and shaded region indicates 95% prediction intervals. Models shown use a bandwidth of 24 months, and drop children eligible for transitional kindergarten (TK).



eFigure 17. Associations Between Elementary School Age-Eligibility and Hospitalizations for COVID-19 by School Period and Model Parameterization. We examined the sensitivity of results various model parametrizations (e.g., local linear regression (loess), linear relationship between age and outcome, and quadratic relationship between age and outcome) and bandwidths. We also considered the effect of excluding individuals born between September and November, as these individuals were eligible for transitional kindergarten (TK). In-person semesters are outlined in black boxes.



eFigure 18. Power Analysis for the Association Between Hospitalizations and School Eligibility. Graph displays the power associated with various effect sizes (incidence rate ratios, IRRs) for the three in-person semesters examined. In our power analysis, we determined that the minimum effect sizes for which we had 80% power were 2.2 (fall 2021), 1.9 (spring 2022), and 3.1 (fall 2022).

eTable 3. Total Number of Cases and Hospitalizations Among the Subsample of Children Who Fell Within 24 Months, in Either Direction, of the Elementary School Attendance Threshold. Bolded rows indicate periods when schools had in-person instruction. For example, for the 2020/2021 academic year, children born before September 1, 2015, would be eligible for kindergarten. Thus, for a bandwidth of 24, children born no more than 24 months before or after September 1, 2015, and who experienced a confirmed COVID-19 infection within the 2020/2021 academic year, would be included in our subsample.

Academic year	Semester investigated	Date ranges	Birthdate cutoff	Total cases	Total hospitalizations
2019/2020	Summer	May 16, 2020 – August 14, 2020	Sept 1, 2014	16,647	87
2020/2021	Fall	August 15, 2020 – December 14, 2020	Sept 1, 2015	51,361	151
2020/2021	Winter	December 15, 2020 – January 14, 2021	Sept 1, 2015	28,650	81
2020/2021	Spring	January 15, 2021 – May 15, 2021	Sept 1, 2015	59,805	170
2020/2021	Summer	May 16, 2021 – August 14, 2021	Sept 1, 2015	36,316	80
2021/2022	**Fall**	August 15, 2021 – December 14, 2021	**Sept 1, 2016**	**134,919**	**183**
2021/2022	Winter	December 15, 2021 – January 14, 2022	Sept 1, 2016	83,327	111
2021/2022	**Spring**	January 15, 2022 – May 15, 2022	**Sept 1, 2016**	**212,093**	**318**
2021/2022	Summer	May 16, 2022 – August 14, 2022	Sept 1, 2016	37,072	121
2022/2023	**Fall**	August 15, 2022 – December 14, 2022	**Sept 1, 2017**	**28,088**	**121**

eTable 4. Incidence Rate Ratios (IRRs) and 95% Confidence Intervals Comparing the Incidence of COVID-19 Among Children Born Just Before the Threshold for Elementary School Attendance (September 1st) Compared to Just After. Rows highlighted in blue are the three in-person semesters included in the study period. Children born between September 2nd and December 1st are eligible for transitional kindergarten (TK), so we conducted analyses both excluding and including children born in September – November. Testing-adjusted estimates attempt to correct for testing differences between schooled and unschooled populations. Results for non in-school periods are the same as in Table 1 in the Main Text.

	Excluding TK-eligible children IRR (95% CI)			**Including TK-eligible children IRR (95% CI)**		
School period	Testing-adjusted: Conservative	Testing-adjusted: Best fit	Not adjusted	Testing-adjusted: Conservative	Testing-adjusted: Best fit	Non adjusted
Fall 2021	1.36 (1.20 - 1.53)	1.52 (1.36 - 1.68)	1.70 (1.54 - 1.87)	1.10 (1.00 - 1.21	1.23 (1.13 - 1.34)	1.38 (1.50 - 1.27)
Spring 2022	1.14 (1.03 - 1.28)	1.26 (1.15 - 1.39)	1.41 (1.29 - 1.54)	0.99 (0.91 - 1.09)	1.11 (1.02 - 1.19)	1.26 (1.16 - 1.36)
Fall 2022	1.06 (0.91 - 1.24)	1.19 (1.03 - 1.38)	1.34 (1.17 - 1.55)	0.94 (0.82 - 1.08)	1.06 (0.94 - 1.21)	1.21 (1.06 - 1.37)

*point-estimates and associated CI's should be interpreted as approximations that attempt to adjust for bias rather than as completely bias-free eTable 5. Results of Meta-Analysis. Values represent the ratio of county-specific incidence rate ratios (representing the association between age-eligibility for elementary school and incidence) for a given unit change in county-level predictor

Variable [source]	**Unit change in predictor**	**Fall 2021-2022** Ratio (95% CI)	**Spring 2021-2022** Ratio (95% CI)	**Fall 2022-2023** Ratio (95% CI)
County population [1]	1 million persons	0.95 (0.91 - 1.00)	0.97 (0.94 - 1.00)	0.94 (0.89 - 1.00)
Population density [1]	1000 persons per square mile	0.96 (0.89 - 1.04)	0.92 (0.86 - 0.99)	1.00 (0.88 - 1.14)
Percentage (%) of people reporting never using a mask [6]	10 percentage points	1.19 (0.68 - 2.08)	1.18 (0.68 - 2.04)	1.09 (0.42 - 2.85)
Vaccination coverage at the start of the semester in children aged 5-11 years [5]	10 additional vaccines per 100 persons	1.05 (0.95 - 1.16)	0.98 (0.93 - 1.04)	1.00 (0.92 - 1.08)
Vaccination coverage at the start of the semester in children aged 12-17 years [5]	10 additional vaccines per 100 persons	1.04 (0.98 - 1.01)	0.97 (0.92 - 1.02)	0.97 (0.90 - 1.06)
Vaccination coverage at the start of the semester in adults aged 18-49 years [5]	10 additional vaccines per 100 persons	1.03 (0.97 - 1.10)	0.96 (0.90 - 1.20)	0.97 (0.90 - 1.06)
Cumulative incidence at the start of the semester (approximation of prior infection) [4]	10 additional cases per 100 persons	0.90 (0.63 - 1.27)	0.88 (0.7 - 1.11)	0.82 (0.6 - 1.12)
Percentage of people living below 150% of the poverty line	10 percentage points	0.94 (0.82 - 1.07)	1.07 (0.94 - 1.21)	1.03 (0.85 - 1.25)
Percentage of people living with disability [3]	10 percentage points	0.89 (0.62 - 1.29)	1.31 (0.93 - 1.87)	0.94 (0.52 - 1.68)
Percentage of families with a single parent [3]	10 percentage points	1.23 (0.71 - 2.15)	1.29 (0.77 - 2.17)	1.40 (0.6 - 3.29)
Percentage of people identifying as an ethnic/racial minority [3]	10 percentage points	1.02 (0.96 - 1.08)	0.96 (0.91 - 1.02)	0.98 (0.9 - 1.08)
Percentage of people identifying as Hispanic [3]	10 percentage points	0.99 (0.93 - 1.05)	1.00 (0.95 - 1.05)	0.98 (0.90 - 1.07)
Percentage of people identifying as Black [3]	10 percentage points	1.15 (0.74 - 1.80)	0.76 (0.49 - 1.18)	1.09 (0.70 - 1.73)
Percentage of people identifying as Asian [3]	10 percentage points	1.04 (0.95 - 1.15)	0.93 (0.86 - 1.02)	1.00 (0.87 - 1.15)
Percentage of people living in crowded housing [3]	10 percentage points	0.92 (0.63 - 1.35)	0.90 (0.65 - 1.26)	0.96 (0.56 - 1.65)

eReferences.

1 U.S. Census Bureau. (2022).
2 Gasparrini, A., Armstrong, B. & Kenward, M. G. Multivariate meta-analysis for non-linear and other multi-parameter associations. *Statistics in medicine* **31**, 3821-3839 (2012). https://doi.org/10.1002/sim.5471
3 Centers for Disease Control and Prevention/ Agency for Toxic Substances and Disease Registry/ Geospatial Research, Analysis, and Services Program. CDC/ATSDR Social Vulnerability Index 2018 Database California. https://www.atsdr.cdc.gov/placeandhealth/svi/data_documentation_download.html. Accessed on Feb 5, 2022.
4 USAFACTS. *Coronavirus Locations: COVID-19 Map by County and State.*, <https://usafacts.org/visualizations/coronavirus-covid-19-spread-map/> (2021).
5 CalHHS. (Sacramento, 2023).
6 Katz, J., Sanger-Katz, M. & Quealy, K. in *The New York Times* (New York, 2020).
7 Hastie, T. & Tibshirani, R. *Generalized Additive Models*. (Chapman and Hall, 1990).
8 Akaike, H. Information theory as an extension of the maximum likelihood principle–In: Second International Symposium on Information Theory (Eds) BN Petrov, F. *Csaki. BNPBF Csaki Budapest: Academiai Kiado* (1973).
9 Symonds, M. R. E. & Moussalli, A. A brief guide to model selection, multimodel inference and model averaging in behavioural ecology using Akaike's information criterion. *Behavioral Ecology and Sociobiology* **65**, 13-21 (2011). https://doi.org/10.1007/s00265-010-1037-6
10 Bilinski, A. & Hatfield, L. A. Nothing to see here? Non-inferiority approaches to parallel trends and other model assumptions. *arXiv preprint arXiv:1805.03273* (2018).
11 California Department of Public Health. *Tracking variants*, <https://www.cdph.ca.gov/Programs/CID/DCDC/Pages/COVID-19/COVID-Variants.aspx> (2021).