Estimating the Effect of Learning Modes in K-12 Schools on COVID Positivity Rates

Final Project Submission for IDS 701: Unifying Data Science

Mohammad Anas, Ying Feng, Vicki Nomwesigwa, Deekshita Saikia (Team 9)

Motivation

The Covid-19 pandemic resulted in a global crisis that impeded everyday lives. As a result, several governments thought it best to set up measures to combat its spread, including closing all schools for several weeks or months. The evidence of reduced influenza incidence rates from the early introduction of restrictive measures during the 2015 to 2016 influenza season in Russia influenced this decision (Ajelli, 2019). Secondly, the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was high in children (Esposito, 2021). Although closing schools is one of the most considerable effective measures to curb Covid -19, its impact is only just emerging. Researchers like Engzell (Verhagen, 2021) and Hammerstein (Frey, 2021) examined the effect Covid-19-related school closures have had on learners' achievement and concluded that students made little to no progress while learning from home, especially students from low socioeconomic status.

While some schools found other ways of continuing to provide education through a mix of online, and hybrid (in-person and remote) learning, others were affected so much that they had to close (Johnson, 2021). As a result, the CDC advised implementing various mitigation measures on schools' operations, especially in communities with increased prevalence of Covid-19 rates. These included but were not limited to the universal correct use of masking, physical distancing, hand hygiene, cleaning, and facility maintenance (including adequate ventilation), and contact tracing with appropriate isolation and quarantine (Prevention, Centers for Disease Control and Prevention. Guidance for COVID-19 Prevention in K-12 Schools., n.d.; Prevention, Overview of Testing for SARS-CoV-2, the virus that causes COVID-19, n.d.). To assess the mitigation measures and monitor the virus's community spread, schools performed repeated screening of all staff and students. Since schools operating in in-person modes would affect how Covid-19 positivity rates change in the community, our research interest is to investigate the causal effect of K-12 schools' different teaching methods on Covid-19 positivity rates in various counties in the U.S.

Data

The datasets used in this research were obtained from *MCH Strategic Data* (Data, 2022) and *The New York Times* (Times, 2022). *MCH Strategic Data* included information for K-12 schools within different school districts in the U.S. with varying methods of teaching and masking policies. This dataset is available for four terms (Fall 2020-2021, Spring 2020-2021, Fall 2021-2022, Spring 2021-2022) for each school district, and had information on student enrollments. Information on Covid-19 positive cases from each county was obtained from *The New York Times*, which records Covid-19 cases within all counties since the beginning of the pandemic in the U.S. To account for the population within each county, we obtained the population information for 2020 from the *United States Census Bureau* (Bureau, 2022).

Since the school dataset was available at a school district level, while Covid-19 cases were recorded at the county level, we aggregated the school dataset to the county level to easily merge with the Covid-19 dataset. The school districts were mapped to counties with the aid of the city-county mapping dataset from *simplemaps* (Simplemaps, 2022).

Since Covid-19 cases were recorded daily, we aggregated to term-level to study the effect of teaching modes per academic term. This results in a panel dataset, aggregated at a county level, for the four academic terms in consideration. The school dataset is merged with the Covid-19 cases dataset with the help of county FIPS codes.

A substantial amount of the records has ‘pending’, ‘other’, or ‘unknown’ reported against their teaching methods and/or student masking policy. To deal with this issue, we consider only those counties where at least 80% of the schools had reported their teaching methods, and this was weighted by the number of school enrollments. Since the trends of Covid-19 positivity rates varied across different time periods in the U.S., owing to surges in different Covid-19 variants, the time effect using the school academic term is also studied.

Summary Statistics

Observing the raw population values, we see that Los Angeles, CA, Cook, IL, and Harris, TX are highly populated counties. Breaking down by academic term, we see the following counties with the highest Covid-19 positivity rates:

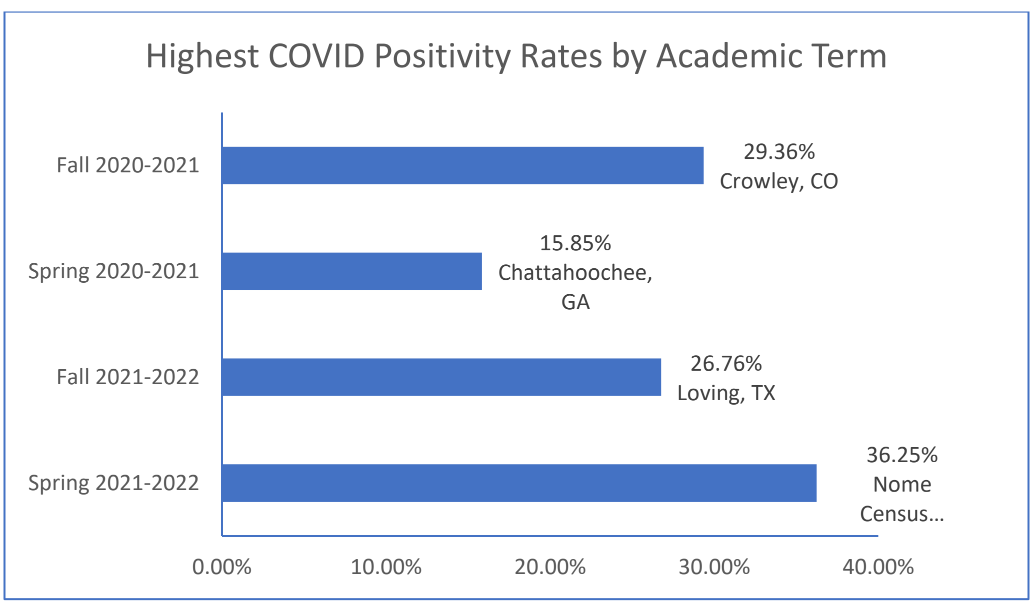

Figure 1:Counties with the highest COVID positivity rates by academic term. Spring 2021 - 2022 academic term saw the highest COVID positivity rates in Nome Census Area County of Alaska.

Analyzing the different kinds of teaching modes adopted by the K-12 schools in the various counties, and corresponding school enrollments, we observe fewer proportions of enrollments in schools which were operating remotely, compared to in-person and hybrid approaches. Because we do not have enough data for remote modes, which could lead to higher standard errors, we carry out this analysis with different definitions of the treatment effect variable, more of which is discussed in the analysis section later.

Term	Full In-Person	Hybrid/Partial	Remote Only
Fall 2020-2021	9.99%	53.76%	36.25%
Spring 2020-2021	38.57%	56.27%	5.15%
Fall 2021-2022	91.82%	8.08%	0.10%
Spring 2021-2022	96.12%	3.69%	0.19%
Grand Total	**57.61%**	**30.85%**	**11.54%**

Table 1: Enrollments in schools in reporting counties by academic term. The highest enrollments in schools in the reporting counties had been observed in schools operating in in-person mode.

The numbers in the table above show that the highest enrollments were observed for schools which were operating in in-person mode. During the school year of 2020 – 2021, there were higher number of students in schools that adopted hybrid approaches, whereas in the following school year, we see higher numbers in schools that adopted in-person teaching.

Model

We note that due to many unobserved factors across counties like testing facilities and isolation laws, the Covid-19 positivity rates vary across counties. This variation can be seen in the plot below, which shows the variation in Covid-19 positivity rates across counties in the U.S.

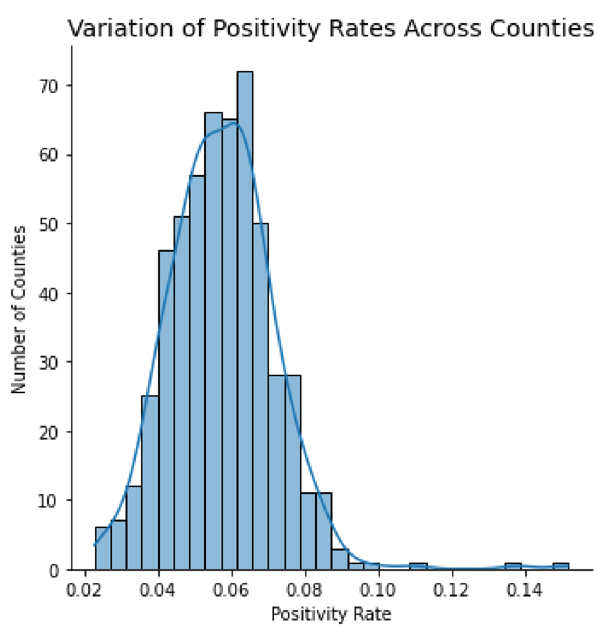

Figure 2: Density plot for Covid-19 positivity rates across all counties.

The variation in Covid-19 positivity rates was also observed across time due to reasons like several variants of Covid-19, and varying vaccination rates across the U.S. over different time periods. The variation in the positivity rates across time can be seen in Figure 3 below.

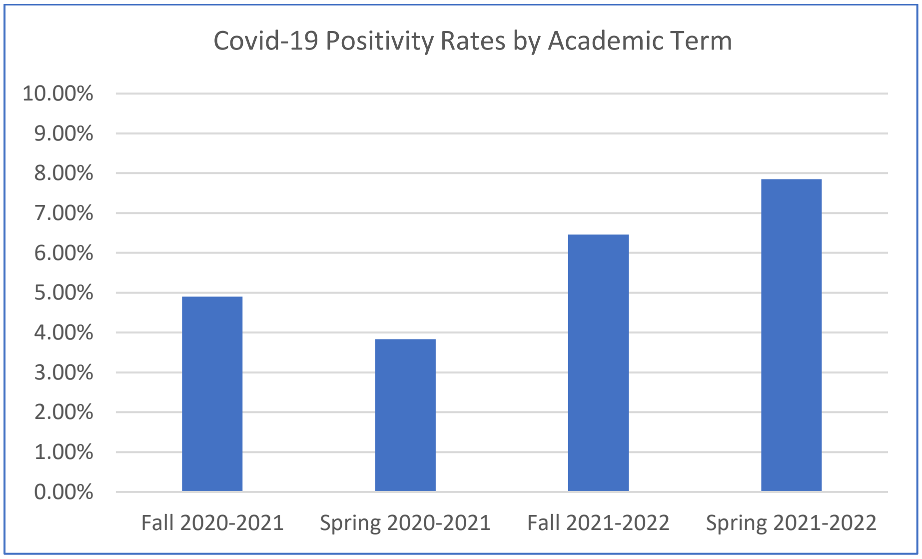

Figure 3: Aggregate Covid-19 Positivity Rates by Academic Term

Given that this variation was explained by factors that were either unavailable in our data or unobserved, a two-way fixed effects model allowed us to address some sources that could possibly lead to omitted variable bias. By controlling for counties as entities, we could eliminate baseline differences among counties. Controlling for school academic terms, we could control for all factors varying across time throughout all counties.

Analysis

As noted above, the amount of data that we have for enrollments in schools where the teaching mode was remote is significantly low as compared to hybrid or in-person teaching modes, therefore, to ensure that we get robust results, we create three variations of a treatment variable, and separate panel data regressions were run for each variation.

The first variation of treatment considered all three teaching modes. A numerical attribute was created in our data, which was assigned the value of 0 if remote teaching was adopted in a particular school, 1 if hybrid teaching was adopted, and 2 if full in-person attendance was required. This treatment attribute was then aggregated across counties and school term by calculating weighted averages by school enrollments.

The second variation of our treatment variable was computed in a similar manner as the previous one. However, this time the schools where hybrid or in person teaching mode was offered were mapped as 1, and schools where the teaching mode was remote were mapped as 0.

For the third variation in the treatment variable, a categorical variable was created which indicated the most dominant teaching mode in terms of enrollments in a county during a particular school term.

The Covid-19 positivity rates for each county was our outcome variable of interest. The results of our fixed-effects models, using the three variations of the treatment variable, are as shown below. The standard errors reported below are clustered across entities.

	Coefficient	**Value**	**Std. Error**	**T-statistic**	**P-Value**
Model 1	Variation 1	0.0045	0.0022	2.0139	***0.0442***
Model 2	Variation 2	0.0097	0.0045	2.1524	***0.0315***
Model 3	Remote	-0.0054	0.0033	-1.6429	***0.1006***
	In-Person	0.0014	0.0019	0.7135	***0.4756***

Table 2: Fixed-effects models summary, with variations of the treatment variable.

For models 1 and 2, we observe that as teaching modes become more inclined towards in-person learning, the Covid-19 positivity rates increase. The p-values indicate this is a significant effect. For the third model, we observed that in counties where most students were enrolled in schools that adopted remote teaching, the Covid-19 positivity rate was found to be 0.0054 units lower than in counties where hybrid was the most dominant teaching mode in terms of enrollments. The low p-value indicates evidence of a significant effect. Moreover, we found that in counties where most schools had adopted in-person teaching, the Covid-19 positivity rates did not differ much compared to counties where hybrid teaching was the most dominant method in terms of enrollments. The difference on average was found to be 0.0014, which was insignificant, as observed by the corresponding p-value in the table above.

Limitations and Conclusion

One major limitation of our model is that the proportions of different categories in our treatment variable are severely imbalanced. To counter this, additional counties that were previously dropped from our analysis were considered, but the proportions remained the same. The number of schools adopting remote teaching modes was significantly less than the number of schools that adopted hybrid and in-person teaching modes. One potential reason for this could be the early studies on Covid-19 cases, which suggested that children could not spread the disease.

From our model, we can conclude that adopting in-person teaching modes during the pandemic led to a rise in Covid-19 positivity rates. This effect was significant if the teaching mode was in-person or hybrid compared to remote teaching. However, the increase in the Covid-19 positivity rates caused by in-person teaching modes was insignificant compared to the causal effect of hybrid teaching modes. Our results align with the recent studies on Covid-19, which suggest that children can also spread the infection, especially for later variants.

References

1. Ajelli, M. L.-H. (2019). Reactive school closure weakens the network of social interactions and reduces the spread of influenza. *The Proceedings of the National Academy of Sciences, 116*(27), 13174-13181.
2. Bureau, U. S. (2022 , April 7). *Index of /programs-surveys/popest/tables/2020-2021/counties/totals. (n.d.)*. Retrieved from https://www2.census.gov/programs-surveys/popest/tables/2020-2021/counties/totals/
3. Data, M. S. (2022, April 4). *COVID-19 IMPACT: School District Operation Status*. Retrieved from https://www.mchdata.com/covid19/schoolclosings
4. Esposito, S. C. (2021). Comprehensive and safe school strategy during COVID-19 pandemic. *Ital J Pediatr*, *47*(6).
5. Frey, S. H. (2021). Effects of COVID-19-Related School Closures on Student Achievement-A Systematic Review. *Frontiers in Psychology, 12*.
6. Johnson, C. S. (2021). Impact of Covid-19 on Educational Change: Back to School. *Springer*.
7. Prevention, C. f. (n.d.). *Centers for Disease Control and Prevention. Guidance for COVID-19 Prevention in K-12 Schools.* Retrieved April 07, 2022, from https://www.cdc.gov/coronavirus/2019-ncov/community/schools-childcare/k-12-guidance.html
8. Prevention, C. f. (n.d.). *Overview of Testing for SARS-CoV-2, the virus that causes COVID-19*. Retrieved April 07, 2022, from https://www.cdc.gov/coronavirus/2019-ncov/hcp/testing-overview.html?CDC_AA_refVal=https%3A%2F%2Fwww.cdc.gov%2Fcoronavirus%2F2019-ncov%2Fphp%2Ftesting%2Fexpanded-screening-testing.html
9. Simplemaps. (2022, April 7). *Database, United States Cities*. Retrieved from simplemaps: https://simplemaps.com/data/us-cities
10. Times, N. Y. (2022, April 4). *Github Repository*. Retrieved from https://raw.githubusercontent.com/nytimes/covid-19-data/master/us-counties.csv
11. Verhagen, P. E. (2021). Learning loss due to school closures during the COVID-19 pandemic. *PNAS, 118*(17).