The Effect of Multigenerational Households on School Instructional Modes and Hospitalizations during COVID-19

Ganqi (Thomas) Li
Econ 80: Advanced Topics in Econometrics
Professor Douglas Staiger
November 2024

Abstract

I examine the effect of multigenerational households on school instructional mode choices and COVID-19 hospitalization rates across U.S. counties during the 2020-2021 school year. I find counties with higher shares of multigenerational households tend to have higher hospitalization rates prior to the start of the school year, and are more likely to select virtual schooling, and that the in-person school mode is associated with a 40% increase in hospitalization rates relative to the virtual school mode. However, conditional on selecting into a school mode, there does not seem to be any observable effect on hospitalization rates due to the interaction between school mode and the share of multigenerational households across counties.

Acknowledgments

I am very grateful to Prof. Douglas Staiger for pointing me to key datasets and offering insightful feedback on my progress, to Prof. Chris Snyder for igniting my interest in health and pandemic outcomes, to Prof. Bruce Sacerdote for showing me relevant literature on education, to Michael (Chengguo) Zhang for sharing his code for cleaning hospitalization data, and to all my friends and mentors who accompanied me throughout my college journey and time after time encouraged me to not give up on myself.

I. Introduction

The COVID-19 pandemic has created unprecedented challenges for schools, families, and communities. School instructional mode decisions (virtual, hybrid, or in-person) have varied significantly across counties, reflecting differing priorities and local circumstances. These decisions are especially critical for multigenerational households, where older adults and school-aged children may share living spaces, which may heighten the risk of severe health outcomes from COVID-19. Using variation in school mode and share of multigenerational households across counties, I investigate the effect of both variables on COVID-19 hospitalization trends, and find that while a higher share of multigenerational households plays a role in counties' decisions to embrace virtual learning, the variation in household structure across counties do not play an additional role in determining hospitalization trends conditional on schooling selection.

II. Literature Review

Ertem et al. (2021) show regional variation in the impact of school modes on COVID-19 cases, so while virtual school modes may widen academic gaps (Goldhaber et al., 2023), perhaps the return in lowered transmission rates is worth the price. I study how the share of multigenerational households affect the decision for virtual schooling, and I follow Ertem et al.'s specification to identify the impact of school modes and the share of multigenerational households on COVID-19 hospitalizations instead of cases, since older adults are more likely to be hospitalized from COVID-19 infection through the shared living spaces with their grandchildren, assuming in-person serves as a mechanism that exacerbates the spread of the disease.

III. Theoretical Framework

Under a simple SIR (Susceptible-Infectious-Recovered) model, the population experiencing an disease outbreak is divided into three groups: those who are not yet infected but susceptible (S), those who are currently infected and capable of transmitting the disease to the susceptible group (I), and those who have recovered from the disease and are no long susceptible because they have gained immunity (R). The dynamics of the model are characterized by

$$\frac{dS}{dt} = -\beta \frac{SI}{N}, \qquad \frac{dI}{dt} = \beta \frac{SI}{N} - \gamma I, \qquad \frac{dR}{dt} = \gamma I,$$

where β is the transmission rate and γ is the recovery rate, and the population is $S + I + R = N$.

Let h be the constant share of newly infected individuals getting hospitalized, and let $H_t = h I_t$ be the number of hospitalized individuals at time t. I would have

$$\frac{dH_t}{dt} \quad = \quad h\left(\frac{dI_t}{dt}\right) \quad = \quad \beta \frac{S_t H_t}{N} - \gamma H_t .$$

If I assume $S_t \approx N$, which admittedly is only true during the early stage of an epidemic, I get

$$\frac{dH_t}{dt} \approx \beta H_t - \gamma H_t = (\beta - \gamma) H_t \qquad \Longrightarrow \qquad \% \, \Delta H_{t+1} \approx (\beta - \gamma),$$

so if I also assume the recovery rate γ is similar across counties, I can assess the difference in transmission rates by examining the difference in percent change in COVID-19 hospitalizations that is associated with school modes and multigenerational household shares.

IV. Data

I obtain data on weekly suspected and confirmed COVID-19 hospitalizations at a hospital level from HealthData.gov, monthly shares of school mode (virtual, hybrid, or in-person) at a school district level from the COVID-19 School Data Hub (CSDH), and individual level demographics from the ACS 2016-2020 5-year sample. I use urbanicity from Geocorr and county level mobility patterns from Google's COVID-19 Community Mobility Reports as controls. The Google data shows changes in mobility, such as in retail workplace activities, relative to a baseline mobility level, defined as the median value from the 5-week period of January 3 to February 6, 2020. I display more details on used datasets in Table 1.

Since the school mode data is on a monthly basis, I assume the school term starts within the first week of the first month that a given school district's data is available, and I only keep school districts that started school in either August or September of 2020. While the monthly dominant school mode within each county may vary across time, I first find the dominant school mode each week, and then find the dominant school mode across all weeks within the school year to assign each county an overall dominant school mode. I merge the variables of interest into a weekly-county level panel dataset, where I only retain counties with available NCES school districts from CSDH and PUMAs from ACS that represent more than 50% of the county's population. Following Ertem et al. (2021), I subset my data to be between 5 weeks before and 13 weeks after the start of school, and prior to the end of 2020, for my empirical specifications, which leaves me with 1,855 counties where hospitalization, multigenerational household share and school mode are all available during the first week of school. Summary statistics for the first week of school are displayed in Table 2.

Table 1—Data Sources

This table displays the datasets and crosswalks for constructing the weekly-county level panel dataset used for this paper.

Dataset	Purpose	Unit	Geography	Frequency	Timeframe	Source
COVID-19 Reported Patient Impact and Hospital Capacity by Facility	Hospitalization count	Hospital	FIPS county	Weekly	2019/12/29-2024/04/21	HealthData.gov
County Population in 2020 Census	Standardize hospitalization rate	County	FIPS County	-	-	Geocorr
ACS 2016-2020 5-Year Sample	Multigenerational households	Individual	PUMA	-	-	IPUMS
District-Monthly Percentage In-Person, Hybrid, or Virtual	Instructional mode	School district	NCES district	Monthly	2020/08-2021/06	COVID-19 School Data Hub
Community Mobility Report	Control	County	FIPS county	Daily	2020/02/15-2020/12/31	Google
Urban-Rural Portion by County	Control	County	FIPS county	-	-	Geocorr
FIPS to PUMA Crosswalk	-	-	-	-	-	Geocorr
FIPS to NCES District Crosswalk	-	-	-	-	-	Geocorr
FIPS to State Crosswalk	-	-	-	-	-	Geocorr
Merged Dataset	-	County	FIPS county	Weekly	2020/06-2021/05	-

Table 2—Summary Statistics with Counties at Start of School Year

This table displays summary statistics for COVID-19 hospitalization, county characteristics and school mode at the start of school year, or at 0 week from start of school, within each county.

	Observations	Mean	SD	Min	Max
Hospitalization (per 100K)	1,855	8.19	17.1	0	246
Share urban (p.p.)	2,699	35.9	33.7	0	100
Share multigenerational (p.p.)	1,855	4.23	0.725	2.06	12.1
Share ages 50+ (p.p.)	1,855	37.4	2.41	18.0	54.3
Share ages 5 to 18 (p.p.)	1,855	17.8	0.846	12.3	26.5
Share virtual (p.p.)	2,699	52.0	45.4	0	100
Share hybrid (p.p.)	2,699	28.4	39.6	0	100
Share in-person (p.p.)	2,699	19.6	36.4	0	100

V. Empirical Framework

To see if county characteristics predict school mode choices, I first run a logit regression of an indicator for a predominately virtual school mode week at the start of school year within each county on county demographics and urbanicity. I then proceed to examine the effect of school mode and multigenerational households on hospitalization outcomes.

Following Ertem et al. (2021), I assess the marginal effect of school mode on percent change of hospitalization using the following Poisson specification with robust standard errors,

$$\ln(Y_{it}) = \pi + \sum \eta \text{Mode}_i + \sum \theta \, \text{Mode}_i \times \text{Week}_t \qquad (1)$$
$$+ \sum \text{State}_i \times \text{Week}_t + \gamma X_{it} + \alpha_i + \omega_t + \epsilon_{it},$$

where Y_{it} is the hospitalization rate, or hospitalization count per 100,000 residents, for county i in week t. I control for county urbanicity and mobility with the vector of controls X_{it}, and I also use county fixed effect α_i and week fixed effect ω_t. I use $\sum \eta \text{Mode}_i$ to control for existing differences in hospitalization rates across counties with differing dominant school modes, and I use $\sum \text{State}_i \times \text{Week}_t$ to control for within-state trends potentially induced by policy. My coefficient of interest is θ, which captures the weekly difference between in-person and virtual and the weekly difference between hybrid and virtual school modes.

I then add interaction terms between county-dominant school mode and the share of multigenerational households to Equation (1) to assess their combined marginal effect,

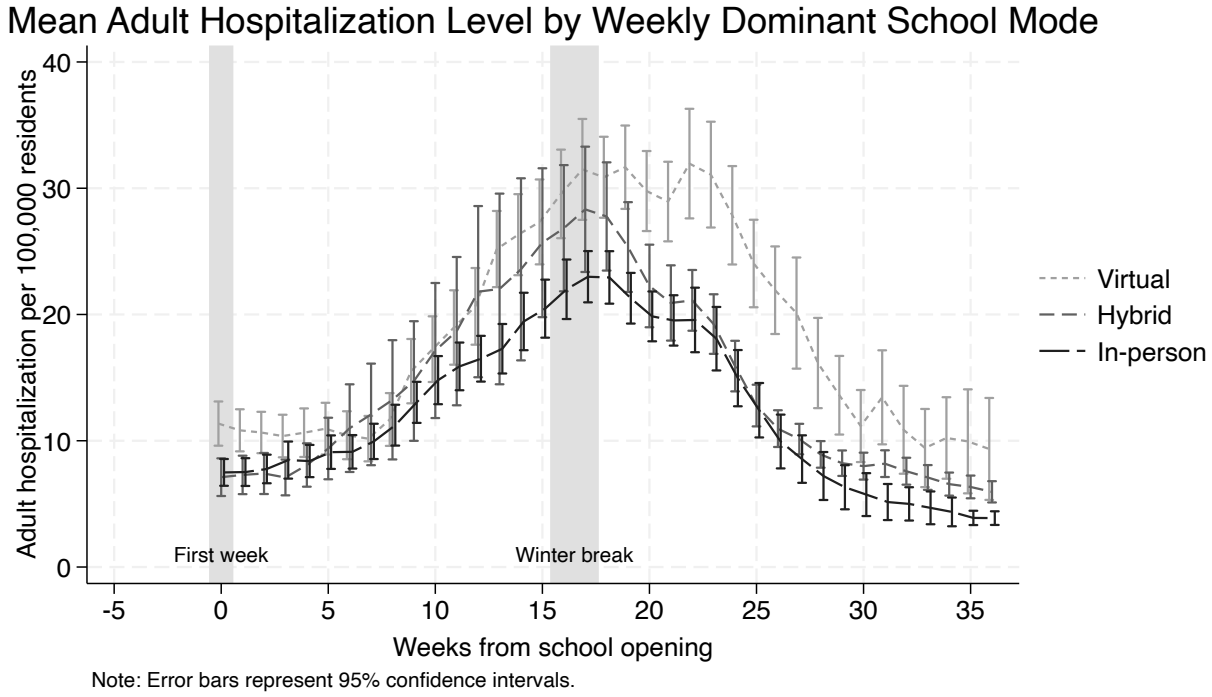
$$\ln(Y_{it}) = \pi + \sum \eta \text{Mode}_i + \sum \theta \, \text{Mode}_i \times \text{Week}_t \qquad (2)$$
$$+ \sum \text{Mode}_i \times \text{Week}_t \times \text{MultigenShare}_i$$
$$+ \sum \text{State}_i \times \text{Week}_t + \gamma X_{it} + \alpha_i + \omega_t + \epsilon_{it}.$$

VI. Empirical Results

I first graph trends in hospitalization rates across counties with different weekly dominant school modes in Table 1. Weekly-varying school modes are available at the start of the school year, defined as 0 week from school opening, and are capped at 36 weeks from school opening. Upon entering the school year, counties with virtual instructions have a higher rate of hospitalization than hybrid and in-person mode counties. Similarly, after winter break, counties that select into virtual schooling have a higher rate of hospitalization. It appears counties choosing the in-person school mode have the lowest mean hospitalization rate throughout most of the school year.

Figure 1—Hospitalization by Weekly Dominant School Mode within Counties

This table displays hospitalization trends across counties with different weekly dominant school modes. Since weekly-level school modes are only available after the school year starts, the trend lines start at 0 week from school opening. Since monthly school mode data is used, the weekly school modes during winter break, or weeks 16 to 17, are likely assigned based upon the school mode preceding the break.

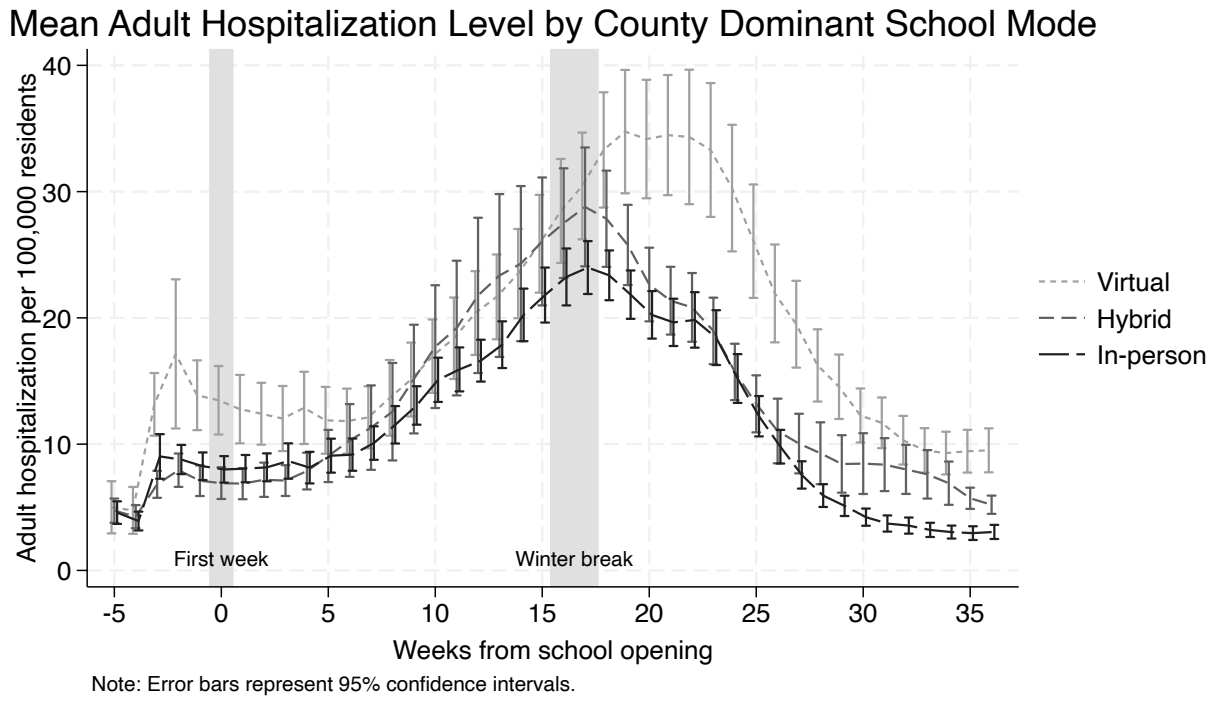


Note: Error bars represent 95% confidence intervals.

I then assign counties into each of the three instructional modes based upon their dominant mode across the entire 2020 to 2021 school year, and graph hospitalization trends in Figure 2. Since each county has a corresponding dominant school mode across the school year, I plot pre-school-year trends from -5 to -1 weeks from school opening across school modes. A similar trend holds, where counties with predominantly virtual schooling also observe higher hospitalization rates, especially at the start of school after the summer and winter breaks.

Figure 2—Hospitalization by Dominant Mode during School Year within County

This table displays hospitalizations trends across counties with different dominant school modes across the school year. Since each county is assigned a dominant school mode, trends in hospitalization can be assigned to different dominant school modes prior to the start of school, unlike in Figure 1. The winter break is from week 16 to 17.



A. Multigenerational Households and School Mode

Since counties self-select into school modes, I look for the effect of county characteristics on the decision to conduct schooling virtually. In the correlation matrix in Table 3, it appears the decision to conduct virtual schooling is correlated positively with the share of multigenerational households and the hospitalization rate at the start of school year. Interestingly, the share of multigenerational households is negatively correlated with the share of population ages 50+, which is helpful in identifying whether the greater share of older population or the greater share of multigenerational households with older adults is the key driver behind virtual schooling.

Table 3—Correlation Matrix at the Start of School Year

This table displays the relationship between the share of multigenerational households, the share of county population ages 50+, the share of county population ages 5 to 18, the number of COVID-19 adult hospitalizations, and the indicator for a virtual-mode dominant week across counties during week 0, or the start of the 2020-2021 school year. While 2,699 counties have available school mode data, only 1,855 have available data on demographic characteristics and hospitalization counts.

	ACS multigen share	Ages 50+ share	Ages 5-18 share	Week 0 adult hospitalization	Week 0 virtual
ACS multigen share	1.0000				
Ages 50+ share	-0.3892	1.0000			
Ages 5-18 share	0.4639	-0.5561	1.0000		
Week 0 adult hospitalization	0.1180	-0.0866	0.0574	1.0000	
Week 0 virtual	0.1258	-0.1452	0.0166	0.0947	1.0000

In Table 4, I run logit regressions of the decision to conduct virtual school during the first week of the school year on county characteristics, and I find the probability of selecting a virtual mode of schooling increases with the proportion of multigenerational households as well as

urbanicity, which represents the share of county population that resides in urban areas. Both findings are reasonable, since more multigenerational families and higher population density might make the effect of in-person learning on COVID transmissions more worrying. However, the probability of virtual schooling also decreases with both the share of ages 50+ and the share of ages 5-18 within counties, even though the shares of these two age groups have opposite signs of correlations with the share of multigenerational households.

Table 4—Logit Regression of Virtual School Mode on County Demographics

This table displays logit regression results of an indicator variable for a virtual-mode dominant week on county household and age characteristics across counties at the start of the 2020-2021 school year.

	Week 0 dominant virtual school mode (1=virtual)	
ACS multigen share (p.p.)	0.377***	0.265***
	(0.0733)	(0.0861)
Share ages 50+ (p.p.)		-0.0994***
		(0.0285)
Share ages 5 to 18 (p.p.)		-0.231***
		(0.0810)
Share urban (p.p.)		0.0136***
		(0.00198)
Observations	1,855	1,855

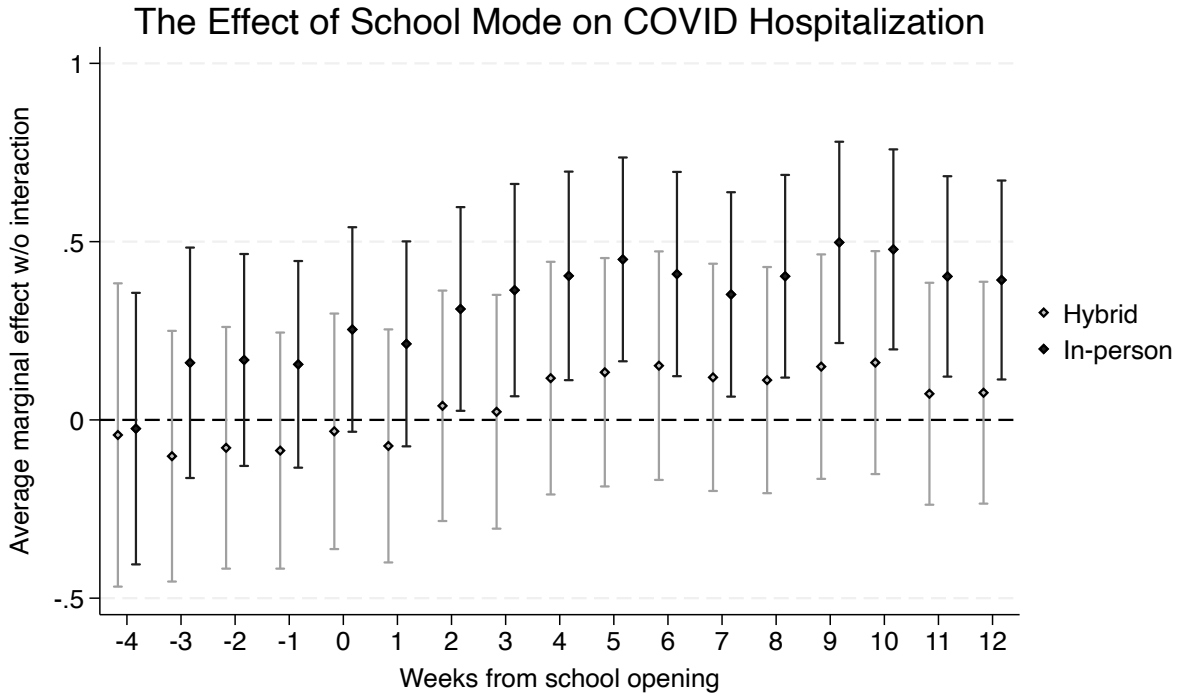
I also provide a graph on the confidence intervals for the share of multigenerational households by dominant school mode in Figure A1 in the Appendix, but more rigorous investigation is required in determining the selection mechanism into school modes.

B. The Relationship between School Mode and Hospitalization

I follow Ertem et al. (2021) in focusing on -4 to 12 weeks from school opening for my Poisson regression specifications, since Google mobility data is only available for 2020. I do not run the regressions by regions (Northeast, Midwest, West, and South) as they do, since the regressions did not converge for all regions. The magical effect of county dominant school on the percent difference in hospitalization rate under Equation (1) is displayed in Figure 3.

Figure 3—The Effect of School Mode on Hospitalization from Eq. (1)

This figure displays the marginal effects of hybrid and in-person county-dominant school mode on the percent difference in COVID-19 hospitalization rates each week relative to virtual schooling with 95% confidence intervals. An effect of 0.1 represents a 10% difference in hospitalization rate as a result of school mode.



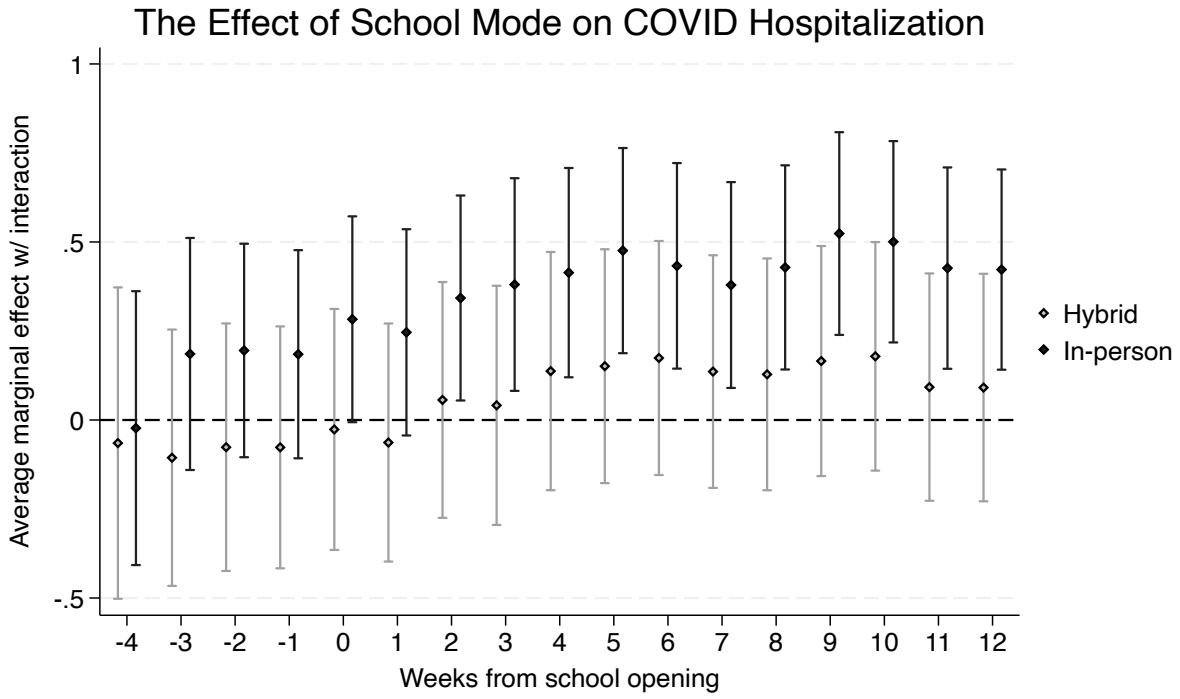
Overall, the marginal effect of hybrid mode relative to the virtual mode on hospitalization rates is not significantly different than 0, but the marginal effect of in-person mode relative to the virtual mode is statistically significant at the 5% significance level, and the marginal effect appears to produce a 40% higher hospitalization rate starting in 2 weeks from school opening.

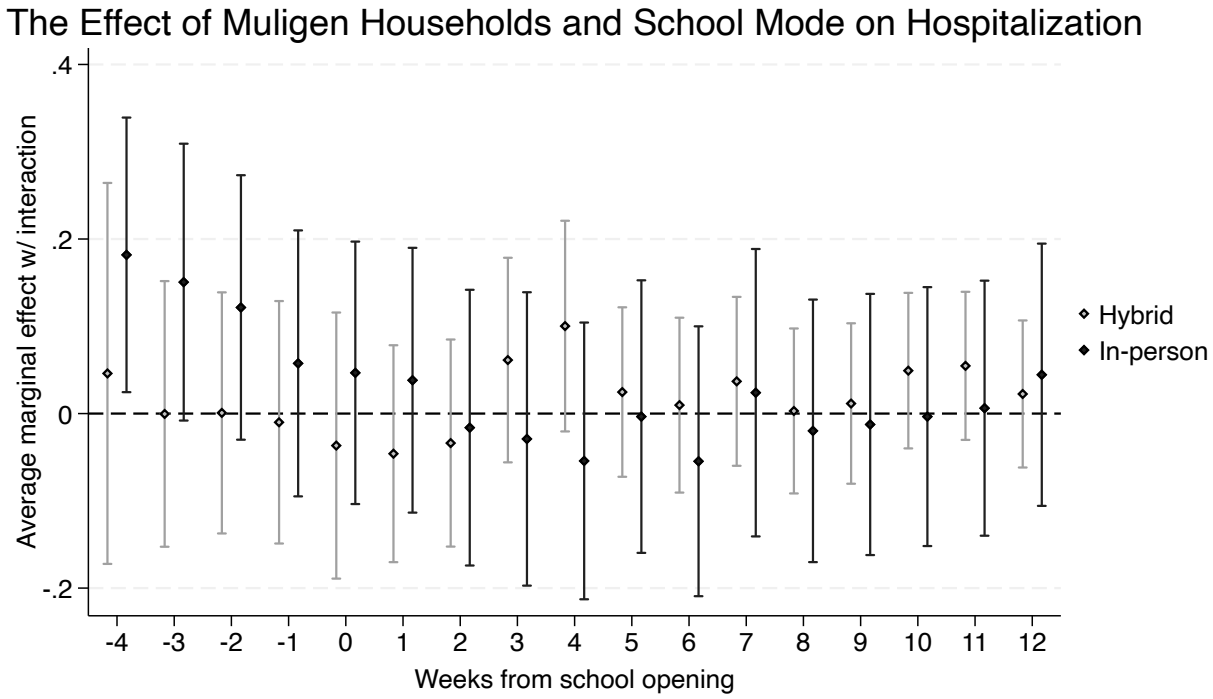
C. The Effect of Multigenerational Households through School Mode

Using regression Equation (2), I plot the marginal effect of school mode and the marginal effect of school mode interacted with the share of multigenerational households in Figure 4. It appears the effect of instructional mode on hospitalization rates is fairly consistent across Figures 3 and 4 at around a 40% difference starting in 2 weeks from school opening, while the variations in multigenerational households fail to explain for any additional statistically significant difference in hospitalization rates after the start of school. Perhaps conditional on using the share of multigenerational households to self-select into a particular dominant school mode, there is insufficient variation in multigenerational shares to produce a noticeable difference in hospitalization rates within each school mode.

Figure 4—The Effect of School Mode on Hospitalization from Eq. (2)

The first graph plots the marginal effect of school modes, and the second plots the marginal effect of interactions.





VII. Conclusion

In this paper, I show that counties with higher shares of multigenerational households were more likely to select virtual schooling, which likely reflects concerns over the vulnerability of older adults in shared living spaces. While in-person schooling is associated with a statistically significant increase in hospitalization rates relative to virtual school, approximately 40% higher starting in two weeks after reopening, the interaction effects between school mode and household composition did not yield additional significant impacts, suggesting that self-selection into instructional modes may account for much of the observed variation in hospitalization trends.

While I use the dominant school mode across the entire school year for each county, I doubt using the weekly-varying county mode would yield a very different result. Nonetheless, more work is required to pinpoint the source of hospitalization rate variation across counties that self-selected into different school modes.

References

Cooper, I., Mondal, A., & Antonopoulos, C. G. (2020). A SIR model assumption for the spread of COVID-19 in different communities. Chaos, Solitons & Fractals, 139, 110057. https://doi.org/10.1016/j.chaos.2020.110057

Ellison, G. (2024). Implications of heterogeneous SIR models for analyses of COVID-19. Review of Economic Design. https://doi.org/10.1007/s10058-024-00355-z

Ertem, Z., Schechter-Perkins, E. M., Oster, E., van den Berg, P., Epshtein, I., Chaiyakunapruk, N., Wilson, F. A., Perencevich, E., Pettey, W. B. P., Branch-Elliman, W., & Nelson, R. E. (2021). The impact of school opening model on SARS-CoV-2 community incidence and mortality. Nature Medicine, 27(12), 2120–2126. https://doi.org/10.1038/s41591-021-01563-8

Fox, S. J., Lachmann, M., Tec, M., Pasco, R., Woody, S., Du, Z., Wang, X., Ingle, T., Javan, E., Dahan, M., Gaither, K., Escott, M. E. A., Adler, S. I., Johnston, S. C., Scott, J. G., & Meyers, L. A. (2022). Real-time pandemic surveillance using hospital admissions and mobility data. PNAS, 119(7). https://doi.org/10.1073/pnas.2111870119

Goldhaber, D., Kane, T. J., McEachin, A., Morton, E. A., Patterson, T., & Staiger, D. O. (2023). The Educational Consequences of Remote and Hybrid Instruction during the Pandemic. American Economic Review: Insights, 5(3), 377–392. https://doi.org/10.1257/aeri.20220180

Lessler, J., Grabowski, M. K., Grantz, K. H., Badillo-Goicoechea, E., Metcalf, C. J. E., Lupton-Smith, C., Azman, A. S., & Stuart, E. A. (2021). Household COVID-19 risk and in-person schooling. Science, 372(6546). https://doi.org/10.1126/science.abh2939

Appendix

Figure A1—Share of Multigenerational Households by County Dominant School Mode

This table displays the mean and 95% confidence interval of the share of multigenerational households and the share of households with both ages 5 to 18 and 65+ across dominant school modes within the school year.

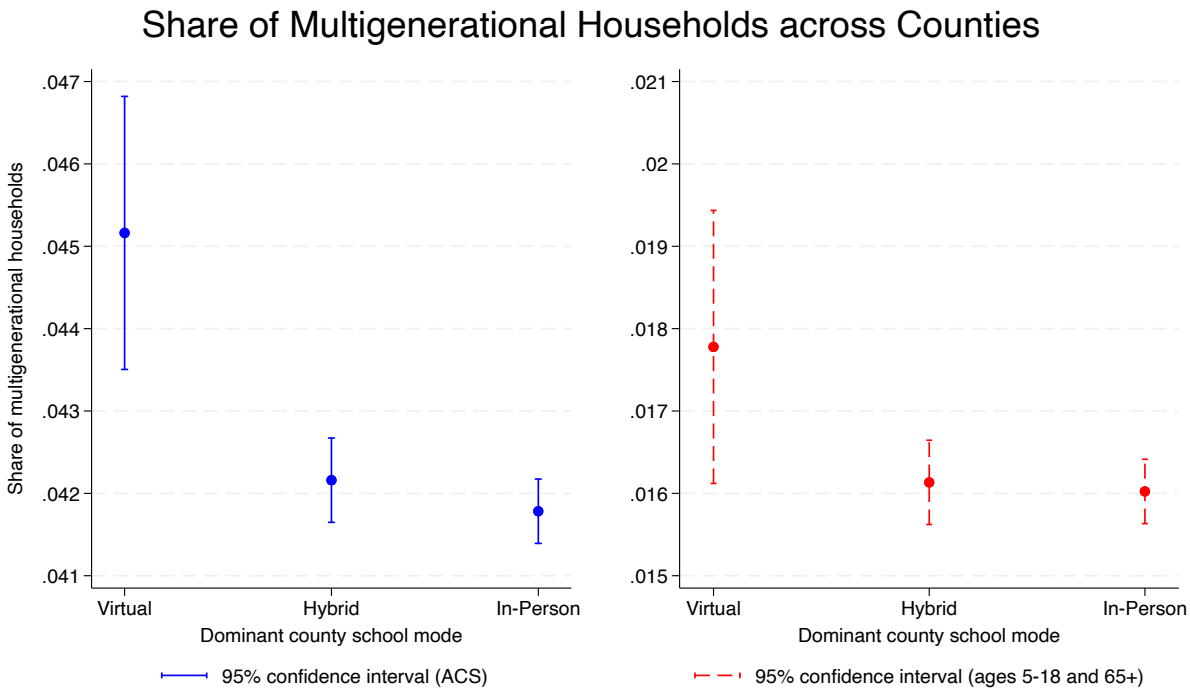


Figure A2—Weekly Hospitalization by Multigenerational Share Quartiles

This table displays hospitalization trends for counties sorted into quartiles by their multigenerational shares.