Association of University Reopening Policies with New Confirmed COVID-19 Cases in the United States

Yang Li[a†], Cheng Ma[a†], Weijing Tang[a†], Xuefei Zhang[a†], Ji Zhu[a] ,Brahmajee K. Nallamothu[b*] (0000-0003-4331-6649).

a. Department of Statistics, University of Michigan, Ann Arbor, MI, 48109, USA.
b. School of Medicine, University of Michigan, Ann Arbor, MI, 48109, USA.

*Brahmajee K. Nallamothu, 132-W, B16, 2800 Plymouth Road, Ann Arbor, Michigan 48109.

Phone Number: (734)647-1624

Email: bnallamo@med.umich.edu

Author Contributions:

Acquisition, analysis, or interpretation of data; Li, Ma, Tang, Zhang, Zhu, Nallamothu

Statistical analysis: Li, Ma, Tang, Zhang, Zhu,

Drafting the manuscript: Li, Ma, Tang, Zhang

Critical revision of the manuscript for important intellectual content: Li, Ma, Tang, Zhang, Zhu, Nallamothu

Supervision: Zhu, Nallamothu

Competing Interest Statement: The authors have declared that no competing interest exist.

Classification: Major: Biological sciences; Minor: Medical Sciences

Keywords: infectious disease; COVID-19; statistical modeling; public policy

This PDF file includes: Main Text, Figures 1 to 4, Tables 1 to 2

†: These authors contributed equally to this work.

Abstract

Reopening of universities in the U.S. has been controversial in the setting of the coronavirus disease 2019 (COVID-19) pandemic. We leveraged several publicly available data sources to study the association of county-level new confirmed COVID-19 case rates since September 1st and the number of students returning to campus across 2,893 U.S. counties with and without universities. In 1,069 U.S. counties with universities, we also studied the association of different reopening policies (online, in-person, hybrid) on new confirmed COVID-19 cases. Multivariate regression models estimated both effects of university reopening and different reopening policies. Mean number of daily confirmed cases per 10,000 county population was 1.51 from August 1st to August 31st, and 1.98 from September 1st to October 22nd. Mean number of students returning to universities was 2.1% (95% CI, 1.8% to 2.3%) of the county population and the number of students returning to campus had a positive association ($\beta = 2.006$, p-value < 0.001) with new confirmed COVID-19 cases within the local county region where the institution resided. For U.S. counties with universities, the mean proportion of online enrollment within each county was 40.1% (95% CI, 37.4% to 42.8%), with most students enrolling in-person or hybrid mode. In comparison to holding class in-person, reopening universities online ($\beta = -0.329$, p-value < 0.001) or in a hybrid mode ($\beta = -0.272$, p-value $= 0.012$) was negatively associated with new confirmed COVID-19 cases. These findings could help public health officials consider policies to mitigate additional waves of infection during the upcoming winter.

Significance Statement

Our study finds that higher numbers of students returning to campus was associated with an increase in new confirmed COVID-19 cases; reopening online or partially online was associated with slower spread of the virus, in comparison to in-person reopening. These findings could provide guidance for policymakers on universities' reopening in upcoming semesters.

Main Text

Introduction

The new academic year for universities and colleges in the U.S. began in the midst of the COVID-19 pandemic in late August and early September 2020. During March and April, most schools moved their classes fully online and locked down campuses. For the 2020 fall semester, many schools opted to reopen fully or partially, given a desire to optimize learning environments and the challenges of virtual classrooms. However, other schools chose to institute policies supporting an online or hybrid mode of learning given concerns for student safety. Though all schools have tried to take many public health actions to ensure reopening occurs safely, numerous reports in the lay media suggest universities and colleges experienced outbreaks of COVID-19 cases after students returned that have extended to surrounding communities (1,2).

Despite such anecdotal observations, the precise effects of students returning to campus on their local regions remain uncertain. Furthermore, how specific university and college reopening policies may be linked to the spread of coronavirus is still unclear. In this study, we investigate these concerns more specifically. We focus both on the association between the spread of coronavirus since September 1st with the number of students returning to campus in each county across the U.S. as well as how different reopening policies are correlated with new confirmed COVID-19 cases.

Materials and Methods

Study Design

Our study had 2 key aims. First, we looked at all counties in the U.S. to study the association between the spread of coronavirus from August 1st to August 31st with the number of students returning to campus in each county. In this aim of the study, we did not distinguish between different reopening policies of universities and colleges (except for excluding those with an online mode of reopening only) and primarily focus on the proportion of population enrollment in universities and colleges in the local county region to understand how the number of students returning to campus may be related to new confirmed COVID-19 cases more generally in their community. Second, we narrowed our focus to the U.S. counties with universities and colleges to specifically study the association of different reopening policies and new confirmed COVID-19 cases. Using publicly available information (3, 4), we divided different policies into three categories: in-person, hybrid, and online. We explored the extent to which virtual classes through hybrid or online reopening – rather than in-person classes – was associated with new confirmed COVID-19 cases.

Data Source

We aggregated multiple publicly available datasets for these analyses. First, we used county-level daily new confirmed cases during the period of August 1st to October 22nd, 2020 (the specific dates of our study time period) from the COVID-19 tracking project (5) by 1Point3Acres (6) which was last updated on October 23rd, 2020. Second, we also obtained state-level cumulative testing rates from the CDC COVID Data Tracker (7). Third, we collected various university and college reopening policies and data on their enrollment from 2,958 institutions across 1,253 counties utilizing resources from the Chronicle of Higher Education (3) which was last updated on October 1st, 2020. (Note: when we use the term university in isolation below we refer to both universities and colleges of higher education after secondary school.) These data were originally provided by the College Crisis Initiative at Davidson College (4). Fourth, we obtained county-level demographics data and land area from the United States Census Bureau (8) website and socioeconomic data from the Economic Research Service in United States Department of Agriculture (9). In particular, we collected the county-level resident population estimates by age

and race in 2019, the county-level median household income in 2018, and the county land area in square miles in 2010.

Outcome Variables

For both aims, our main outcome variable was the mean number of daily new confirmed cases per 10,000 county population from September 1st to October 22nd at the county level. Note that we chose to use the data starting from September 1st to capture the trend of confirmed cases after reopening, because the majority of these reopened near late August or early September. Moreover, we considered the mean number of daily new confirmed cases over a relatively long period (until the end date of analysis on October 22nd, 2020), rather than shorter intervals, to smooth out the weekly periodic pattern and the lag in impact of various reopening policies. In addition, we standardized the county-level average number of daily new confirmed cases by the county population as the total population varies significantly over different counties and it may affect the absolute number of confirmed cases for each county.

Exposure Variables

We used different exposure variables for the two aims as their goals are different. To study the impact of universities' reopening, we focused on the number of students enrolled through in-person or hybrid classes trying to examine the effect of reopening between counties with and without universities. To further compare different reopening policies, we focused on specific reopening policies in only counties with universities. The original dataset for reopening policies mainly categorized five policies including "fully in person", "primarily in person", "hybrid", "primarily online", and "fully online". For ease of interpretation, we further merge "fully in person" and "primarily in person" as in-person policy, and "fully online" and "primarily online" as online policy, which resulted in three policy categories: in-person, online, and hybrid in our analysis. Below we describe the exposure variables used for the two aims in detail.

To study the impact of universities' reopening, we used an exposure variable that captured the number of university students returning to their campus within each county. Given that the enrollment and policies of universities in each county may differ when more than one university is present in a county, we defined the exposure variable for this county as follows. We first aggregated all universities in a county and summed up the enrollment of students under the in-person or hybrid policies (where students would likely return to campus). Then the exposure variable was obtained as the ratio of this sum and the total county population (*ENROLLP*). Note that we used the total enrollment of universities under the above two policies as a reasonable proxy of the number of university students returning to the campus. And, as there is a large variance among universities' enrollment, the proposed exposure variable better characterizes the population flow, in comparison to alternatives such as the number of universities with in-person or hybrid policies. A higher value of the exposure variable means that more students stayed in physical buildings and were involved in onsite campus activities per county population. The exposure variable equals zero if there was no university in a region (or all universities in a region took online reopening policy).

To compare different reopening policies across universities, we focused on a subset of 1,069 counties with universities. The main exposure variables were the proportion of enrollment for three policies within each county (*ONLINE%*, *HYBRID%*, *INPERSON%*). For example, for each county, the "in-person" variable was the ratio of the total enrollment of all universities taking in-person policy in this county to the total county enrollment. The sum of three policy exposure variables equals to one. We were interested in whether the proportion of hybrid and online enrollment would be negatively associated with the COVID-19 confirmed cases when taking in-person enrollment as the baseline. We showed an example of this calculation in Figure 4 for Washtenaw County where the University of Michigan is one of several schools in the region.

Other Control Variables

We considered the following control variables in models for both aims. We first included the mean number of daily new confirmed COVID-19 cases per 10,000 county population from August 1st to August 31st (*AUGP*). Examining new confirmed COVID-19 cases in August before reopening occurred would allow us to account for baseline differences in the severity of coronavirus infections across counties. We also controlled for five county-level demographic and economic characteristics that could potentially be associated with new confirmed COVID-19 cases including: 1) the proportion of population aged over 65 years (*AGE65*), 2) the proportion of population aged below 20 years (*AGE20*), 3) the proportion of population of racial minorities (non-White) (*RACEMINO*), 4) the median household income as of 2018 (*MEDHHINC*), and 5) COVID-19 testing rates at the state level (*TR*). Note that for the variable testing rates, we only had data available at the state level and used this as a proxy for county-level resources for testing.

Statistical Analysis

The analysis was done at the county level. The dataset initially provided county-level daily confirmed cases for 3,086 counties (out of 3,141 counties in the U.S.). To focus on in-person, online, and hybrid policies of interest, we excluded 163 counties that contained any university with "other" or "undetermined" policies. We observed a handful of "negative" daily confirmed cases, which were used to correct historical mistakes in the county-level data (i.e., a case was re-categorized the following day from being positive to negative). Given that such data from counties with large numbers of corrections are likely to be unstable and may even indicate limited surveillance systems, we excluded 30 (~1%) counties that contained at least one day with more than 5 corrected cases, which resulted in 2,893 counties in our analysis.

We used multivariate linear regression models to estimate the effects of university reopening and different reopening policies. Note that the enrollment proportion of three policies sum up to 1, so we dropped the covariate in-person enrollment proportion to avoid the issue of collinearity among covariates. We used a p-value of less than 0.05 to indicate statistically significant evidence of an association between the covariate and the outcome variable.

Lastly, we conducted three sensitivity analyses.

First, we conducted another analysis to compare the early-stage (within a month) and late-stage (after a month) impact of university reopening. Specifically, for each county, we considered two periods of time, one in September and the other in October. Thus, each county has two outcome measurements, corresponding to the mean number of daily new confirmed cases per 10,000 county population in September and in October respectively. We also constructed an indicator variable indicating whether the period under consideration is October (*I(Oct)*), and added itself and its interactions with the mean daily new confirmed proportion in August and the enrollment proportion into the linear regression model to account for the variation over time.

Second, we changed the exposure variable - the in-person and hybrid enrollment proportion - into a binary variable. This variable is 1 if the in-person and hybrid enrollment proportion is greater than zero, and is 0 otherwise. The goal of this sensitivity analysis was to determine if a "dose-response" effect was noted based on the proportion of in-person and hybrid enrollment (rather than its mere presence in a county).

Third, because the impact of universities' reopening to counties with large populations can be relatively minor, we further removed 87 counties with the top 5% total county population and re-estimated the model based on this cohort.

Results

Cohort Characteristics

Our final study cohort included 2,893 U.S. counties with a subset of 1,069 counties where at least one university or college was present. Characteristics of these U.S. counties are provided in Table 1. The mean of daily confirmed cases per 10,000 county population was 1.51 from August 1st to August 31st, and 1.98 from September 1st to October 22nd. The overall total number of daily new confirmed cases across the U.S. shows an increasing trend starting from early September (Figure 1). Mean number of students returning to schools was 2.1% (95% CI, 1.8% to 2.3%) of the county population. For those counties with universities, the proportion of online enrollment within each county is on average 40.1% (95% CI, 37.4% to 42.8%), with the rest of the students enrolling in-person or hybrid mode.

Students Returning to Campus and New Confirmed COVID-19 Cases

Mean number of daily new confirmed COVID-19 cases per 10,000 county population from June 1st to October 22nd over U.S. counties stratified by the presence of returning students is plotted in Figure 2. Mean number of daily new confirmed COVID-19 cases in U.S. counties with the top 50% in-person or hybrid university enrollment rates increased at a higher rate compared with U.S. counties without in-person or hybrid universities. This separation appeared largest during an early phase in September with infections rising in U.S. counties without in-person or hybrid colleges in October.

The result of our fitted model after multivariable adjustment is shown in Table 2. Counties with higher proportions of in-person or hybrid university enrollment were associated with higher confirmed case rates since September ($\beta = 2.006$, p-value < 0.001). That means, for every 10% rise in the ratio of in-person or hybrid university enrollment to the county population, the mean daily new confirmed cases would increase by about 0.2 persons per 10,000 county population (e.g., 41.8 cases per 30 days in a median-size U.S. county where at least one university was present).

Additional variables beyond students returning to campus were associated with new confirmed COVID-19 cases as well. For example, the average confirmed case rate in August was positively correlated with the confirmed case rate after September ($\beta = 0.204$, p-value < 0.001). The proportion of young people under the age of 20 and the proportion of older adults 65 years or older also were both positively associated with the outcome (p-values < 0.001). This could be due to the fact that older adults are at higher risk of getting COVID-19, and young people would likely be going back to school after September, bringing higher chances of COVID-19 spread. Median household income, which can in general represent the local economic status, was negatively correlated with the outcome ($\beta = -0.015$, p-value < 0.001) suggesting lower numbers of new confirmed COVID-19 cases in wealthier regions. As expected, the regional testing rate was positively associated with the outcome ($\beta = 1.810$, p-value < 0.001). Finally, we have found that the proportion of population of racial minorities in a county was negatively correlated with the outcome ($\beta = -0.955$, p-value <0.001). This result was robust using other control variables that represent racial minorities (e.g., African American proportion in a population). This may be due to our evaluation of new confirmed cases late in the pandemic (i.e., time after September), instead of the whole pandemic period where rates of infection have been higher in U.S. counties with higher proportions of minority populations.

Reopening Policies and New Confirmed COVID-19 Cases

Figure 3 shows the mean number of daily confirmed COVID-19 cases per 10,000 population over 318 U.S. counties with all universities taking an in-person policy, 173 U.S. counties with all universities taking a hybrid policy, and 303 U.S. counties with all universities taking an online policy. Daily confirmed cases per 10,000 population of three groups were similar until the end of August, but after that, there were gaps between 3 groups (as reopening of schools occurred).

Among 1,069 counties with universities, we estimated models after multivariable adjustment and displayed these results in Table 3. Online enrollment proportion was negatively associated with the outcome (β = -0.329, p-value < 0.001). This result could be interpreted in this way: fixing the hybrid enrollment proportion, if we move 10% of the enrolled university students in a county from in-person to online mode, the mean number of daily new confirmed cases per 10,000 county population would have decreased by approximately 6.85 cases per 30 days in a median-size U.S. county where at least one university was present. We also found the hybrid policy enrollment proportion was negatively correlated with the outcome (β = -0.272, p-value = 0.012). This suggests that moving 10% of the enrollment students from in-person mode to hybrid mode would reduce the mean number of daily new confirmed cases per 10,000 county population by approximately 5.67 cases per 30 days in a median-size U.S. county where at least one university was present. The interpretation of other control variables is similar to our earlier results.

Sensitivity Analysis

We performed several sensitivity analyses. We found evidence of an early-period effect with enrollment proportion noted to be a significant variable in affecting the number of new confirmed COVID-19 cases in September, but with its association becoming insignificant in October (p-value = 0.054). This suggests that the effect of school reopening may depend on the time, and the effect is more important in the first few weeks after reopening with a “catch-up” phase in U.S. counties without universities. This time-varying effect is observed in Figure 2, where there was an initial case surge in September with a later rise in cases in October. Secondly, converting the continuous value of enrollment proportion into a binary one leads to weaker association between new confirmed COVID-19 cases and university enrollment. This implies the value of measuring the specific number of in-person and hybrid enrollment as a continuous variable, rather than just a simple indicator, to reflect the extent of university students returning to the campus. Finally, we performed a sensitivity analysis that excluded U.S. counties with large populations to ensure our findings were not susceptible to outliers; these results were consistent with our earlier findings. The details of these model fitting results are summarized in Supplementary Information.

Discussion

In this study, we examined the effect of reopening in universities and colleges on the spread of COVID-19 in their surrounding communities during the fall. Our findings could be summarized in two folds. First, we found that students returning to campus was associated with an increase in new confirmed COVID-19 cases in the counties in which the universities and colleges reside. Second, we assessed the effects of three different reopening policies and found that reopening online or partially online was associated with slower spread of the virus, in comparison to in-person reopening.

Our findings could provide evidence-based guidance on the upcoming winter/spring 2021 semesters. For example, due to the impact of enrollment to its neighboring area, regions with universities and colleges should expect a spike of the confirmed cases in the first few weeks after students return to campus. Medical supplies such as masks and medication should be prepared in advance as well as enforcement of strict social distancing. However, our findings also suggest that these effects wane over time when compared with other U.S. counties without colleges or universities. In addition, for those areas that do choose to reopen, we found evidence to support that an online or hybrid mode with virtual classes may be the preferable option at mitigating the number of new confirmed COVID-19 cases.

These findings have the following limitations. First, we studied the effects of university reopening policies using observational data with multivariate adjustment. The choice of control variables may not be a comprehensive list, and residual confounding from other sources of variability in county-level rates of COVID-19 cases is certainly possible. However, we were able to adjust for several key factors and our results were stable to a few key sensitivity analyses. Second, the

actual reopening dates may vary among universities, ranging from late August to mid-September. In our analysis, the calculation of outcome variable, which is based on the cutoff date September 1st, could be viewed as approximation of confirmed rates after university reopening. Third, testing capacity and how accurate the reported numbers of tests and cases are might vary across counties or over time.

Data Availability

All datasets used are publicly available. Our analysis was conducted using R statistical software version 3.6.0 and 3.6.1. All statistical code is available for replication at https://github.com/yli15/COVID-19-research.

Acknowledgments

This work was done through support and access to the American Heart Association COVID-19 Data Challenge Award and its Precision Medicine Platform. Our team was awarded the final prize as part of this competition. We are grateful to Ms. Jessica Baker with her assistance in preparation of this manuscript.

References

1. S. Hubler, A. Hartocollis, How Colleges Became the New Covid Hot Spots. *The New York Times*, 11 September 2020. https://www.nytimes.com/2020/09/11/us/college-campus-outbreak-covid.html. Accessed 13 October 2020.
2. A. Nierenberg, A.Pasick, A Crackdown in Ann Arbor. *The New York Times*, 21 October 2020. https://www.nytimes.com/2020/10/21/us/university-michigan-stay-in-place-order-testing.html. Accessed 29 October 2020.
3. The Chronicle of Higher Education. Here's Our List of Colleges' Reopening Models. https://www.chronicle.com/article/heres-a-list-of-colleges-plans-for-reopening-in-the-fall/. Accessed 13 October 2020.
4. The College Crisis Initiative. C2i Dashboard. https://collegecrisis.shinyapps.io/dashboard/. Accessed 16 October 2020.
5. T. Yang *et al.*, CovidNet: To Bring the Data Transparency in Era of COVID-19. arXiv:2005.10948 (20 Jul 2020).
6. 1point3acres. Global covid-19 tracker & interactive charts. https://coronavirus.1point3acres.com/. Accessed 23 October 2020.
7. Centers for Disease Control and Prevention. CDC COVID Data Tracker. https://covid.cdc.gov/covid-data-tracker/#testing_testsperformed. Accessed 29 October 2020.
8. Census Bureau. County Population by Characteristics. https://www.census.gov/data/tables/time-series/demo/popest/2010s-counties-detail.html. Accessed 16 October 2020.
9. Economic Research Service. Department of Agriculture. County-level Data Sets.https://www.ers.usda.gov/data-products/county-level-data-sets/. Accessed 16 October 2020.

Figures and Tables

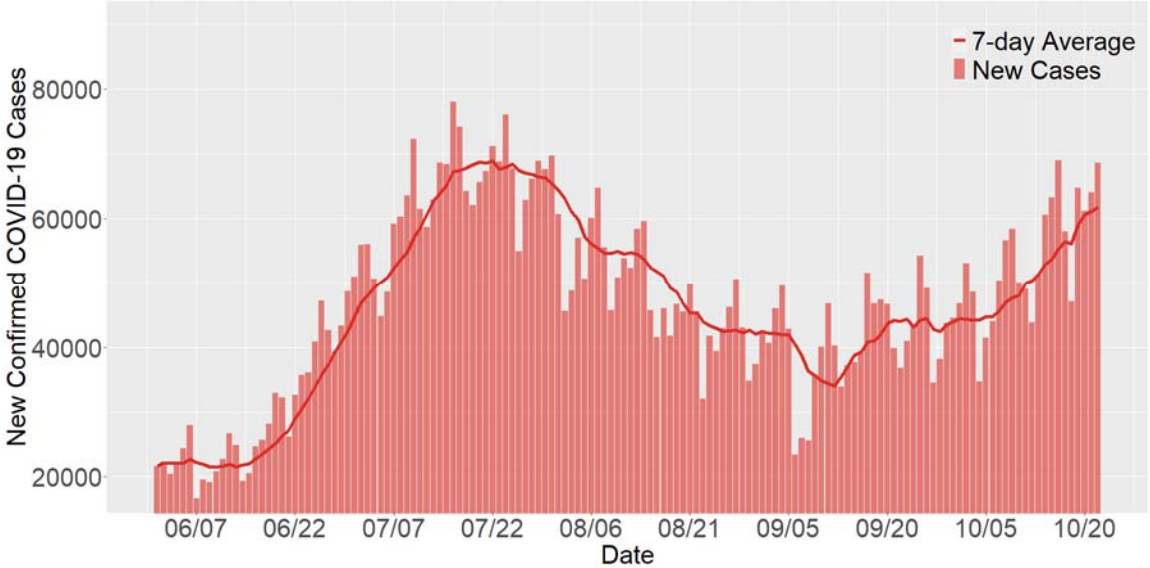

Figure 1. Nationwide total number of daily new confirmed COVID-19 cases since June 1st and their 7-day average. The number of daily new confirmed COVID-19 cases shows an increasing trend starting from early September that coincides with the time of school reopening.

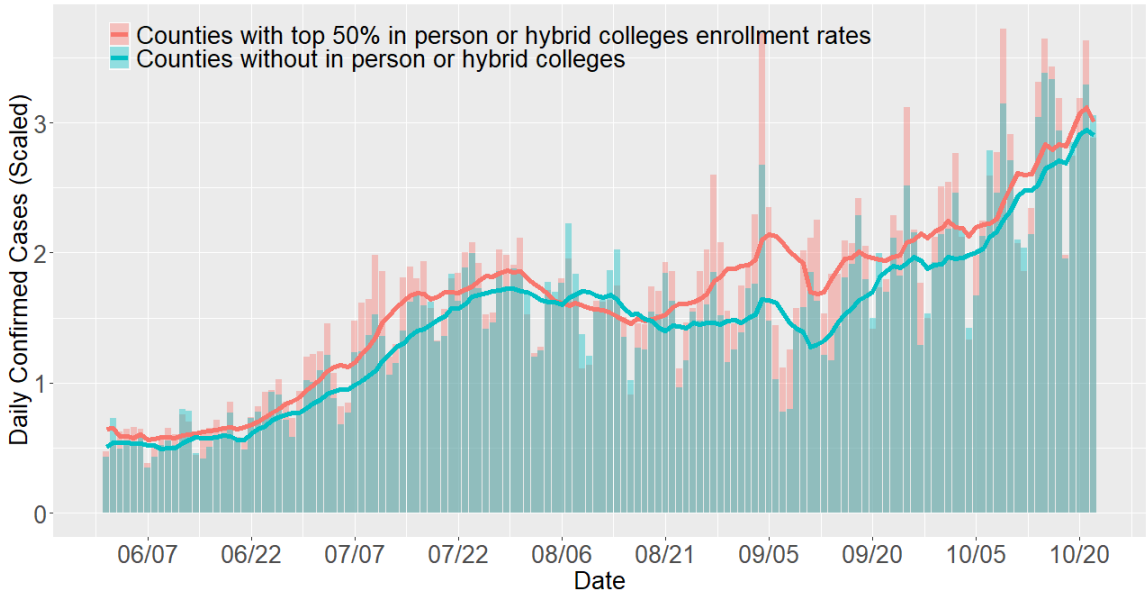

Figure 2. The mean number of daily new confirmed COVID-19 cases per 10,000 county population over U.S. counties without universities with in-person or hybrid reopening policies, and U.S. counties with the top 50% in-person or hybrid university enrollments per 10,000 county population from June 1st to October 22nd. The solid curves represent their 7-day average. The mean number of daily new confirmed COVID-19 cases in U.S. counties with the top 50% in-person or hybrid college enrollment rates appears to increase as compared with U.S. counties without in-person or hybrid colleges. Interestingly, this separation appears to be greatest during an early phase in September with infections rising in U.S. counties without in-person or hybrid colleges in October.

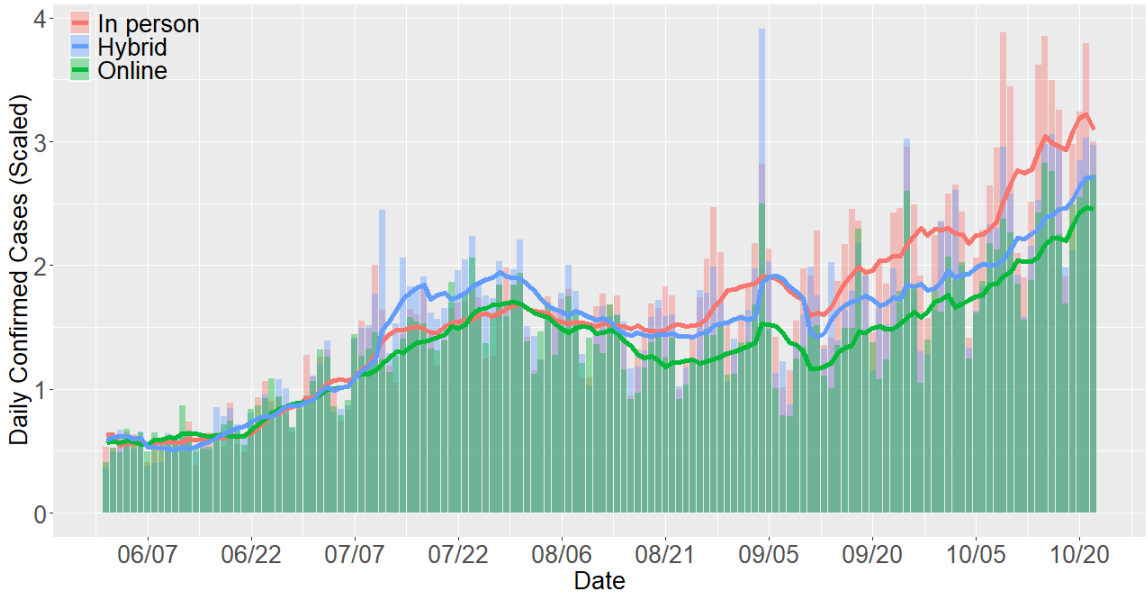

Figure 3. The mean number of daily confirmed COVID-19 cases per 10,000 population over 318 U.S. counties with universities taking an in-person policy, 173 U.S. counties with universities taking a hybrid policy, and 303 U.S. counties with universities taking an online policy. The solid curves represent their 7-day average.

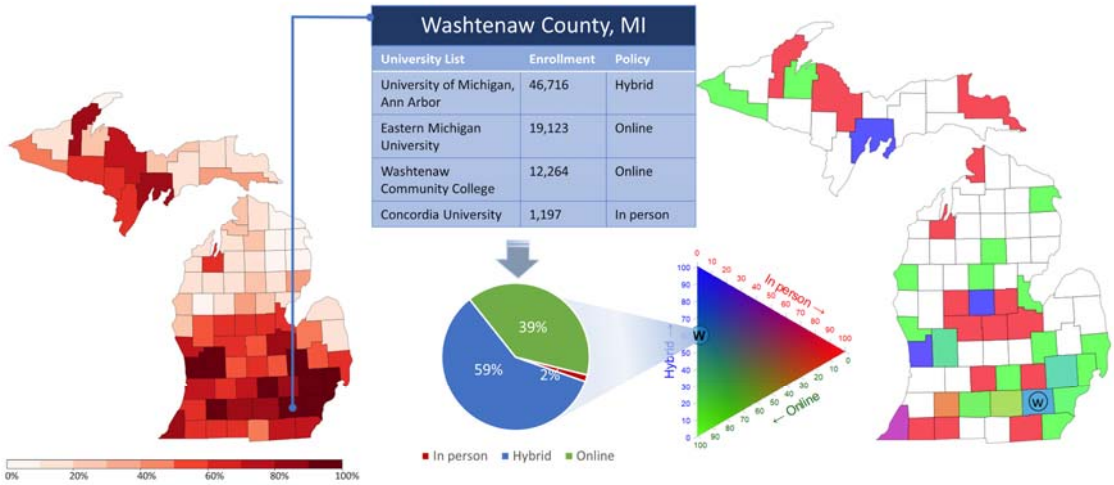

Figure 4. We illustrate how to calculate our exposure variables using our state of Michigan. Left: The percentile rank of the mean daily confirmed cases from September 1st to October 22nd relative to all counties in the U.S. Middle: Schematic illustration of how to construct the main exposure variables in our analysis of U.S. counties with universities and colleges: the proportion of enrollment for three policies within each county. Right: The geographic distribution of universities' reopening policies over counties in Michigan.

Table 1. Cohort Characteristics. The outcome is SEPTP (the average number of daily new confirmed COVID-19 cases per 10,000 county population after September 1st). The main exposure variable is ENROLLP (in-person + hybrid enrollment proportion). The control variables are: AUGP (the average number of daily new confirmed COVID-19 cases per 10,000 county population in August), AGE20 (the proportion of young people under the age of 20), AGE65 (proportion of older adults 65 years or older), RACEMINO (the proportion of population of racial minorities, i.e., non-White), MEDHHINC (the median household income as of 2018), TR (the COVID-19 testing rates at the state level).

	All 2,893 counties under study	**1,069 counties with colleges**
Variables	**Mean(sd)**	**Mean(sd)**
SEPTP	1.98(1.66)	1.87(1.45)
AUGP	1.51(1.78)	1.42(1.17)
ENROLLP	2.05(5.79)	N/A
INPERSON%	N/A	36.1(45.1)
HYBRID%	N/A	23.8(38.8)
ONLINE%	N/A	40.1(45.1)
AGE20	24.3(3.52)	24.8(3.00)
AGE65	19.9(4.70)	18.2(3.82)
RACEMINO	15.1(15.9)	17.6(1.59)
MEDHHINC (thousand USD)	52.2(13.5)	55.0(14.4)
TR	0.413(0.146)	0.422(0.144)

Table 2. The estimated coefficients on 2,893 U.S. counties with or without universities and colleges. The main exposure variable is the sum of in-person and hybrid enrollment proportion (ENROLLP). The control variables are the same as Table 1.

	Estimate	Std Error	P-value
Intercept	-2.507	0.495	<0.001
AUGP	0.204	0.017	<0.001
AGE20	14.148	1.134	<0.001
AGE65	4.373	0.954	<0.001
ENROLLP	2.006	0.508	<0.001
RACEMINO	-0.955	0.199	<0.001
MEDHHINC (thousand USD)	-0.015	0.002	<0.001
TR	1.810	0.201	<0.001

Table 3: The estimated coefficients on 1,069 U.S. counties with universities and colleges. The main exposure variables are policies as enrollment percentages (ONLINE% and HYBRID%). The control variables are the same as Table 1.

	Estimate	Std Error	P-value
Intercept	-0.729	0.743	0.327
AUGP	0.425	0.036	<0.001
AGE20	12.109	1.824	<0.001
AGE65	-1.511	1.499	0.314
ONLINE%	-0.329	0.095	<0.001
HYBRID%	-0.272	0.108	0.012
RACEMINO	-1.454	0.264	<0.001
MEDHHINC (thousand USD)	-0.017	0.003	<0.001
TR	1.581	0.267	<0.001