medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Impacts of teaching modality on U.S. COVID-19 spread in Fall 2020 semester

Syed Badruddoza

Assistant Professor, Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, Texas, United States

Modhurima Dey Amin

Assistant Professor, Department of Agricultural and Applied Economics, Texas Tech University, Lubbock, Texas, United States

Corresponding author: Syed Badruddoza. Agricultural Science 307-A. Texas Tech University. Lubbock, TX 79409. United States. Phone: 309 585 8032. Email: syed.badruddoza@ttu.edu.

Acknowledgments: Authors are grateful for comments and suggestions from the participants of the American Economic Association and Southern Economic Association meetings. There is no conflict of interest. No funding was received for this project.

NOTE: This preprint reports new research that has not been certified by peer review and should not be used to guide clinical practice.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Impacts of teaching modality on U.S. COVID-19 spread in Fall 2020 semester

Abstract

We study the impact of college reopening in Fall 2020 on county-level COVID-19 cases and deaths using the information of 1,076 U.S. colleges. We match college and county characteristics using several methods and calculate the average treatment effects of three teaching modalities: in-person, online, and hybrid on COVID-19 outcomes up to two months after college reopening. In pairwise comparison, colleges reopened with in-person teaching mode were found to have about 36% point more cases within 15 days of reopening, compared to those reopened online, and the gap widens over time at a decreasing rate. Death rates follow the pattern with a time lag. However, colleges with hybrid mode catch the pattern of in-person mode after some time. We also find that greater endowment and student population, and fewer republican votes in the county are major predictors of choosing remote teaching modes over in-person.

Keywords: College, COVID-19, Online, Teaching modality, United States.

JEL codes: A23, I18, I23.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

1 Introduction

The COVID-19 pandemic and its containment measures have unprecedentedly affected human health and economic activities (e.g., McKibbin and Fernando, 2020; Atkeson, 2020). Countries across the world have implemented partial or full business and school closures to mitigate the infection spread (Panovska-Griffiths et al., 2020; Di Domenico et al., 2020). Many U.S. colleges temporarily closed or switched to online in the spring semester, and over six out of ten colleges reopened in the fall semester with an in-person or a combination of in-person and online teaching plans (Gallagher and Palmer, 2020; Picault, 2021). College students mainly fall in the age cohort of 18 to 29 years, which has a lower death rate (0.4%) from COVID-19 (Wrighton and Lawrence, 2020), but a greater chance of socialization than the older age cohorts. Although colleges have taken different measures–including symptom screening, contact tracing, cleaning and disinfecting, and mandatory mask policies–there is a chance that reopening colleges may spread the virus among faculty and staff serving students, and in the region around the college (Hubler and Hartocollis, 2020).

Given the substantial risk of infection spread from college campuses to the community, a timely policy question is, if and to what extent, switching to a more remote teaching modality help in controlling communicable diseases. Although the severity of the pandemic somewhat reduced by the end of 2021, a study that analyzes the effect of different teaching modalities may offer directions in similar crises in the future—such as an expansion of a new variant of COVID-19. The current study applies a quasi-experimental approach to find the impacts of teaching modality on infection spread onto the neighboring areas. Intuitively, we match college and county characteristics to find a pair of similar colleges, but one chose in-person and the other online. Then we measure how choosing an in-person mode might have affected COVID-19 cases and deaths in the neighborhood as opposed

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

to choosing an online mode. We perform similar exercises for two other pairs: in-person versus hybrid, and hybrid versus online. Results provide evidence that college reopening, especially with in-person teaching mode, increases COVID-19 cases and deaths over time.

Most studies so far have looked into the impact of elementary and secondary school closure, optimal reopening plan, and teaching strategy (e.g., Di Domenico et al., 2020; Panovska-Griffiths et al., 2020; Gandolfi, 2020; Cohen et al., 2020). Walke, Honein, and Redfield (2020) discuss the possible issues of U.S. college reopening and the prevention measures required on campuses. In a study close to ours, Andersen et al. (2020) use an event study under difference-in-difference setting to investigate college openings' association with cellphone-tracked human mobility and COVID-19 cases in the counties of the campuses. Their preliminary results suggest that reopening increased county COVID-19 cases by 2.7 per 100 people, of which 0.042 per 100 thousand people resulted in death, and in-person teaching created greater visits and higher cases during the reopening. However, most of these studies do not consider the fact that the choice of reopening and teaching modality is endogenous to observed cases. We use different matching methods to re-weight the sample based on the college's probability of selecting an instructional mode.

2 Method

We are interested in finding the effects of teaching modality on two COVID-19 outcomes: (1) new daily COVID-19 cases reported in the county, and (2) new COVID-19 related deaths in the county. The impact of teaching modality is assessed on county-level outcomes (and not on the college level) because students may spread the virus to non-students and immunocompromised people in the community, which is a critical policy concern. First, we specify a simple relationship between college reopening in the fall semester and the outcome as below,

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

$$Y_i = \text{Intercept} + \alpha \text{T}_i + \text{Controls}_i \times \beta + \text{error}_i \quad (1)$$

where, Y_i is a COVID-19 outcome, $i = 1, \ldots, N$ indicate colleges and T_i is a binary treatment variable that represents teaching modality. We use three specifications of T_i separately: (1) in-person=1, online=0, (2) in-person=1, hybrid=0, and (3) hybrid=1, online=0 for pairwise comparison of teaching modalities. The vector of control variables includes observed college and county level covariates such as college endowments or percent of the population who stayed at home, and β is a vector of respective parameters. Our parameter of interest is α that shows the impact of a mode of instruction on an outcome variable, e.g., whether switching from online to in-person has an impact on the COVID-19 outcomes. The equation above would have been identified if colleges were homogeneous and were randomly assigned a teaching mode (Barnow, Cain, and Goldberger, 1981). In practice, the treatment assignment is not random, because colleges choose the mode of teaching based on their distance education capacities, observed COVID-19 cases, and many other characteristics. The probability of adopting a teaching modality should considerably vary across colleges. Intuitively, if we could select two colleges with identical institutional and regional features, except one went in-person and the other online, we could measure the difference in COVID outcomes induced by the teaching modality.

There are some other estimation issues. First, many areas had high COVID-19 cases in Spring and Summer semesters. Due to the exponential nature of the disease spread, places with more initial cases will experience and faster growth. Moreover, teaching modality in the fall semester may depend on existing cases, thus generating a reverse causality. Second, there can be observed and unobserved college and county features that affect both teaching modality and COVID-19 outcomes, creating an omitted variable bias. Third, a measurement error may occur because infections are reported less during the weekend and more on weekdays, and the incubation or

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

survival period varies. Finally, a county may have multiple colleges that determine the presence of students in the region. We take several measures to address these problems. For instance, we control for a number of observed college and county characteristics, and use percentage changes of the outcome variables to eliminate college and county-level unobserved fixed effects. We also control for the presence of other colleges in a county. Attempts to address endogeneity are described below.

One source of identification is that different colleges reopened on different dates between July and October of 2020. Therefore, we can isolate the effect of reopening on COVID-19 from the aggregate trend. We also implement various matching methods described below. To deal with the measurement and threshold selection issue, we use two-week intervals based on recommended quarantine period: 0-15 days after reopening, 15-30 days after, 30-45 days after, and 45-60 days after reopening, and run separate cross-section regressions for each subsample. The benchmark of percentage calculation is 15-0 days before reopening; for example,

$$\begin{aligned}&\%\text{ change in cases}_{\text{0 to 15 days after reopening}}\\&= 100 \times \frac{\left(\text{cases}_{\text{0 to 15 days after reopening}} - \text{cases}_{\text{0 to 15 days before reopening}}\right)}{\text{cases}_{\text{0 to 15 days before reopening}}}\end{aligned} \tag{2}$$

and so on. We chose two weeks intervals because doing so gives COVID-19 enough time to appear in the outcomes, and we did not go beyond 60 days because the effects may be contaminated by other exogenous factors.

We employ five methods and check if the results are robust: (1) bivariate Ordinary Least Squares (OLS) where $\beta = 0$, (2) multivariate OLS where $\beta \neq 0$, (3) propensity score matching (PSM), (4) Nearest Neighbor matching (NN), (5) and Kernel Multivariate Distance Matching (KMDM). The first two methods are conventional. The latter three methods focus on finding "statistical twins" for each observation in the treatment group from the control group with similar

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

values of the covariates. The average treatment effects can be calculated as the mean of differences between the observed values in the treatment group and the imputed counterfactual values. We present a non-technical version of the matching methods below.

Rosenbaum and Rubin (1983, 1984) discuss propensity score matching (PSM) to adjust the probabilities for the differences in pre-treatment variables. A propensity score is the conditional probability of receiving the treatment given the pretreatment variables, i.e., $P(T_i = 1|X_i)$. The probabilities of treatments generated are used to create weights that adjust the pre-treatment imbalances. For example, if a college has a high probability of choosing online than in-person, greater weights are assigned on in-person than online during the estimation. PSM is usually calculated using a logistic regression (for a recent technical discussion, see Imbens, 2000; McCaffrey et al., 2013).

Although finding a match with PSM is relatively easier, many studies argue that matching with propensity scores can be misleading as it ignores the multidimensional differences between two observations and simplifies them into one dimension—the score (e.g., King and Nielsen, 2019). The nearest neighbor (NN) matching adopts a more multidimensional approach and uses some (default being one) closest observations in the control group. In case of ties, NN uses all ties as matches. Conversely, a single control observation can be used many times with replacement. Matching with more than one continuous covariate may induce bias, so we apply a bias-adjusted NN matching proposed by Abadie and Imbens (2011).

Our final matching technique, MDM, is similar to NN, but uses a certain bandwidth for multivariate matching instead of comparing with the closest neighbor. Simulation studies show that MDM tends to outperform PSM because it approximates fully blocked randomization that is relatively more efficient (King and Nielsen, 2019). We apply KMDM where observations in

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

treatment and control groups are matched using Mahalanobis distance, and use the Epanechnikov kernel function to assign larger weights to controls with smaller distances (see Jann (2017) for further discussion on the advantages of MDM). Following Huber et al. (2015), we choose 1.5 times the 90% quantile of the non-zero distances in finding a pair.

3 Data and descriptive statistics

We collected characteristics of four-year and two-year undergraduate degree-granting colleges and universities randomly chosen from the National Center for Education Statistics (NCES, 2020). Most of the colleges available in the data are public or private non-profit. Data on teaching modality and the official start date of the fall 2020 semester were manually collected between July and December of 2020 by authors from respective college websites[1]. Over 90% of the colleges mentioned their latest teaching modality between June and September 2020. Information for the remaining colleges was obtained from contemporary campus news or local news. Some examples of search phrases we used are, "XXX university fall 2020 covid information reopening plan new" or "XXX college restarts in-person class in fall 2020" or "XXX institute fall 2020 plan president announcement updated". We categorize fall reopening plans into three types: (1) in-person, (2) online, and (3) hybrid. Colleges in the "in-person" group started Fall 2020 semester with face-to-face classes and open residence halls, and may include some online delivery materials; colleges in the "online" group primarily offered online classes with some exceptions for lab components and may have some students on campus; and colleges in the "hybrid" group either divided the class into online and in-person section, or switched the teaching mode on a rolling basis, or offered courses with both in-person and online access. County characteristics were extracted from the American Community

[1] A list of colleges and their webpages are publicly available in school_info.txt on https://github.com.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Survey (ACS, 2020). Merging all these data sets eventually gives 1,076 colleges in all U.S. states and Washington D.C.

We obtain daily new COVID-19 cases and deaths at the county level from the New York Times (NYTimes, 2020). We also merged the data with individual mask policies in county and state from HealthData.gov (2020). Finally, information on county-level shares of republican votes in Presidential Election 2016 is included from the MIT Election Data and Science Lab (MIT, 2018) to take residents' perception of COVID-19 risk into account (e.g., Tyson, 2020).

Figure 1 simply plots the changes in COVID-19 cases and deaths as a percent of 15 pre-semester days grouped by college teaching modality. We took the percentage changes to remove the college and county-level idiosyncratic features. Different colleges started the Fall 2020 semester on different dates from the end of July to October. Figure 1 matches the reopening dates, and suggests that COVID-19 cases were exponentially growing in the Fall semester regardless of teaching modality, and COVID-related deaths followed the pattern. However, colleges teaching online are located in counties that had slower growth of cases after the college reopening date, on average. Cases were slowly growing for colleges with hybrid teaching mode as opposed to colleges with in-person mode. After 30 days, the pattern of cases is reflected in the pattern of deaths. Counties where colleges taught in hybrid or in-person mode experienced a rise in COVID-related deaths, but counties where colleges taught online experienced a fall. However, an econometric analysis is required to find the impact of teaching modality.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

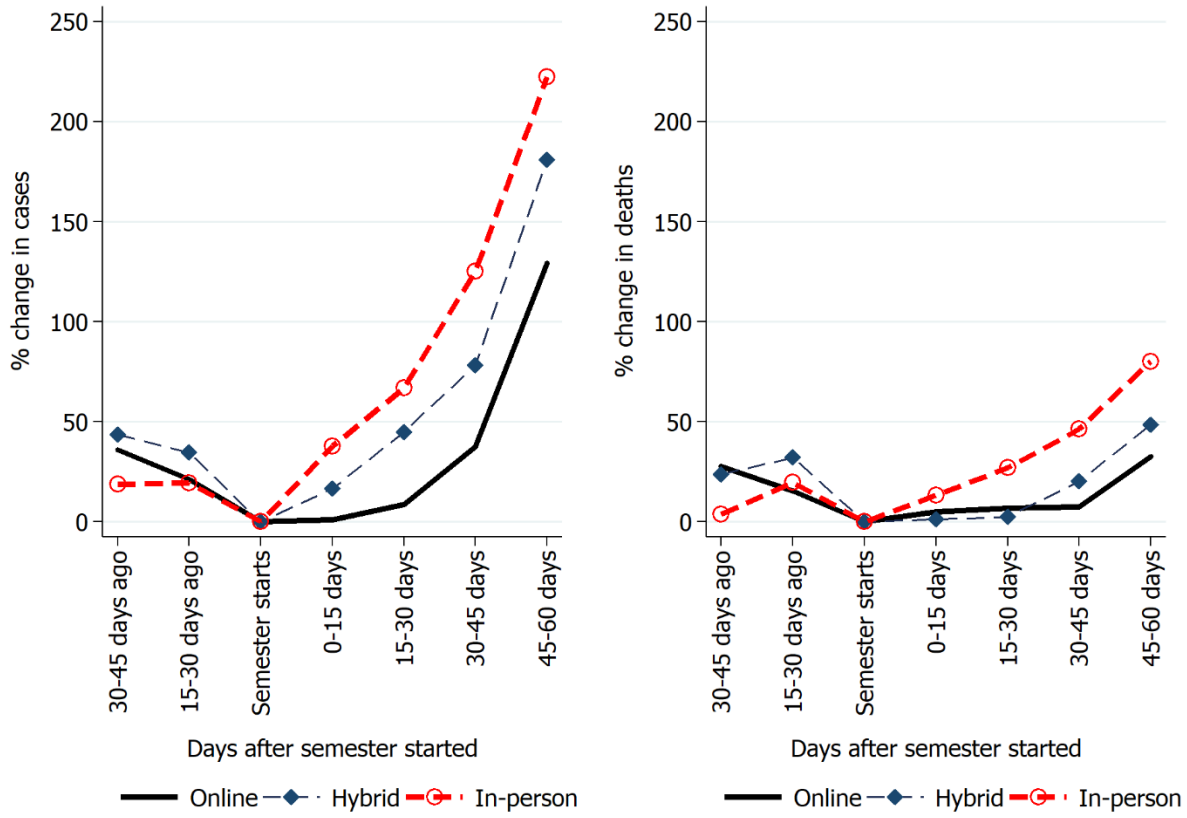


Fig 1. Growth of COVID-19 cases by college teaching modality in Fall 2020 semester.

Note: Figures are averages across colleges. Daily data of 1,076 colleges from July to October 2020. See the data section for details.

Source: Cases are obtained from the New York Times (2020), and teaching modalities were collected by authors from the respective college websites.

Table 1 provides variable descriptions and sources, and Table 2 presents summary statistics of variables. The earliest reopening date for the Fall semester is July 20th in our sample whereas the latest date is October 24th. Our analysis also controls for the average percentage of the county

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

population that stayed home daily, collected from the Bureau of Transportation Statistics (BTS, 2020). Omitting stay-home makes the effects of a teaching mode on COVID-19 outcomes biased, because COVID-19 outcomes may increase if fewer people stay home regardless of campus reopening or teaching modality. A variable on the existing mask ordinance was also included for the same reason. These variables are further used to find the determinants of teaching modality and matching. The matching process should only contain variables that are measured at baseline because variables measured at around the treatment may be influenced or modified by treatment (Austin, 2011). Therefore, we selected the determinants from the latest academic year available before the Fall 2020 semester. We also included total COVID-19 cases and deaths by the end of the Spring 2020 semester because it was the last regular semester that gave the college administration a signal about the disease situation in the community. Some studies find an association between temperature and COVID-19 spread (e.g., Livadiotis, 2020), so the average temperature during the reopening month was included in the model.

A required condition for estimating treatment effects is distributional similarities of pre-treatment covariates across the treatment groups McCaffrey et al. (2013). Table 2 indicates considerable variation across teaching groups. The difference requires the use of matching in our analysis. Tests for the distributional equality of matching variables between the treatment and control group are placed in the appendix.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 1. Variable description.

(1) Variable	(2) Unit	(3) Source	(4) Details
Online	Binary	College websites	Takes 1 if the college holds classes primarily online for Fall 2020, zero otherwise
Hybrid	Binary	College websites	Takes 1 if the college adopts a combination of online and in-person modes for Fall 2020, zero otherwise
In-person	Binary	College websites	Takes 1 if the college holds classes primarily in person for Fall 2020, zero otherwise
New cases	Numeric	NYTimes (2020)	Daily new COVID-19 cases per 100,000 population in the county
New deaths	Numeric	NYTimes (2020)	Daily new COVID-19 deaths per 100,000 population in the county
% stayed home	Percentage	BTS (2020)	Percentage of county population that did not go outside home for longer than 10 minutes
Enrollment	Count	NCES (2020)	Total student count in fall 2018 semester, including graduate and remote students
Cost of attendance	USD	NCES (2020)	Total price for in-state students living on campus for 2019-20
Endowment per student	USD	NCES (2020)	Endowment assets per FTE enrollment at the end of 2018
Student-faculty ratio	Ratio	NCES (2020)	Enrollment divided by total full-time instructional staff
Public	Binary	NCES (2020)	The institution is primarily funded by a state government
Total population	Count	ACS (2020)	Total population in the county where the college is located (2018)
Black population	Percentage	ACS (2020)	Percentage of African American population in the county (2018)
Household income	USD	ACS (2020)	Median real household income in 2018
Republican votes	Percentage	MIT (2018)	Percentage of votes for the Republican Party at the county level in Presidential election 2016
Mask ordinance	Binary	HealthData.gov (2020)	The county has an ongoing mandatory mask policy for individuals in public places
Cases in Spring 2020	Count	NYTimes (2020)	Total COVID-19 cases in the county by May 15, 2020
Deaths in Spring 2020	Count	NYTimes (2020)	Total COVID-19 deaths in the county by May 15, 2020
Temperature	Fahrenheit	NOAA (2021)	County mean temperature in reopening month

Note: Means and standard deviations of the variables are placed in Table 2.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 2. Summary statistics.

(1) Variable	(2) Online	(3) Hybrid	(4) In-person	(5) Overall
Time varying variables				
Daily new cases	247.00	100.04	72.43	145.59
	(558.23)	(282.69)	(248.92)	(408.97)
Daily new deaths	4.12	1.62	1.09	2.38
	(10.54)	(5.98)	(4.39)	(7.75)
Online				0.38
				(0.48)
Hybrid				0.26
				(0.44)
In-person				0.36
				(0.48)
% stayed home	26.41	25.05	24.21	25.27
	(4.06)	(4.33)	(4.00)	(4.22)
Enrollment	8,828.61	6,307.21	4,417.95	6,580.85
	(9,991.36)	(8,272.99)	(6,972.39)	(8,754.75)
Cost of attendance	44,980.83	46,178.28	41,089.31	43,757.31
	(21,312.56)	(19,431.71)	(16,628.42)	(19,059.39)
Endowment per student	61,279.10	28,545.41	59,892.61	51,811.22
	(347,471.23)	(62,887.43)	(367,714.37)	(307,016.55)
Student-faculty ratio	41.71	28.11	29.18	33.61
	(30.23)	(22.71)	(25.53)	(27.44)
Public	0.60	0.38	0.33	0.44
	(0.49)	(0.49)	(0.47)	(0.50)
Total population	1,578,217	616,050	444,525	917,566
	(2,620,932)	(973,588)	(1,028,931)	(1,867,046)
Black population	14.57	13.68	11.89	13.37
	(15.77)	(14.44)	(13.14)	(14.55)
Household income	64,587.54	59,805.99	56,796.75	60,526.88
	(17,710.66)	(15,156.93)	(14,395.72)	(16,253.30)
Republican votes	38.94	46.09	52.97	45.86
	(16.72)	(16.25)	(14.53)	(16.93)
Mask ordinance	0.86	0.80	0.71	0.79
	(0.35)	(0.40)	(0.45)	(0.41)
Cases in Spring 2020	12,706.16	5,356.90	3,576.15	7,498.42
	(23,007.05)	(10,396.52)	(10,177.45)	(16,800.38)
Deaths in Spring 2020	594.02	288.29	70.89	370.25
	(1,038.35)	(568.51)	(5.298)	(789.65)
Temperature (F) in reopening month	72.37	71.38	70.89	71.58
	(6.14)	(6.07)	(5.30)	(5.86)
Observations (Colleges)	406	284	386	1,076

Note: Standard deviations in parentheses. Table 1 provides variable description.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

4 Results

Tables 3 and 4 respectively show the average treatment effects of teaching modalities on COVID-19 cases and deaths. We begin with simple bivariate OLS to test the correlation, then control for other covariates, and then use three matching models. Table 3 suggests that the effect of in-person teaching on COVID-19 cases, as opposed to online, is positive and significant for all time intervals. For example, the bivariate model shows that cases increase by 36.78% point in the first 15 days after reopening, and keep increasing at a decreasing rate in later periods. The hybrid teaching mode also has a positive significant difference from online. The gap between the effects of in-person and online stays when we control for other covariates, but becomes statistically insignificant between hybrid and online. Matching with propensity scores does not indicate a significant advantage of any teaching modality in later periods. A significant effect in later periods is important from a policy perspective because the difference of cases in the first 15 days can be influenced by more tests before reopening with in-person teaching mode. Matching with NN somewhat corroborates the result from the simple OLS. A more robust method like KMDM suggests a statistical difference in COVID-19 cases between in-person and online groups, and to some extent, hybrid and online groups. The gap between in-person and hybrid remains statistically insignificant. To summarize, colleges that chose in-person over online experienced a greater increase in COVID-19 cases in their respective counties. A hybrid instructional mode may create a difference from in-person at the beginning, but may not sustain for longer periods.

Table 4 presents similar results for COVID-19 deaths. MDM captures a statistically significant difference between in-person and online modes, which also grows over time at a decreasing rate. For example, teaching in-person instead of online would result in 33% point more

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

deaths in 30-45 days after reopening compared to the pre-semester 15 days. Other results are not significant at the 1% level, but it is interesting that the gap between hybrid and online groups closes in the 45-60 days interval, while the in-person group remains statistically different from them. As seen earlier, no difference in deaths was found for the hybrid versus online model.

Notice that the magnitude of the average treatment effects increases at a decreasing rate. This occurs because college reopening increases average cases and deaths regardless of teaching modality. Although online mode raises the rates slower than in-person, the gap starts to shrink after 45 days, possibly because college students are likely to socialize on campus outside the classroom even if the classes are online.

The dataset further allows us to explore another aspect of teaching modality during a pandemic. We use logistic regressions to find the predictors of a teaching modality. The results in Table 5 show that colleges with greater enrollment are less likely to choose in-person than online (column 1), and in-person than hybrid (column 3), holding other variables constant. A percent increase in endowment per student decreases the log odds ratio of choosing in-person over online by 0.28, and that of choosing hybrid over online by 0.26, given the other predictors. Public colleges are less likely to choose an in-person mode as opposed to online, but the coefficient is significant only at the 10% level. A greater population in the county might have made colleges choose some remote teaching components. Prevailing mask ordinance has a negative association with going in-person, and more republican votes have a positive association with choosing in-person over hybrid or online. A negative association of in-person mode with deaths in the Spring 2020 semester is noticeable, but not statistically significant. However, there is some positive association between the percentage of the Black population in the county and choosing an in-person mode over hybrid. The variable representing the percentage of people who stayed home is not statistically significant in

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

these results, but it allows us to control for the mobility in the area along with controlling for the total population. To summarize, greater enrollment, endowment per student, county population, fewer republican supporters in the county, and prevailing individual mask ordinance are major predictors of adopting online or hybrid teaching modality over in-person.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 3. Average treatment effects on COVID-19 cases

(Dependent variable: % change in cases compared to cases in two weeks before reopening date)

		(1)	(2)	(3)	(4)
Model	Y variable	0-15 Days	15-30 Days	30-45 Days	45-60 Days
1. Biv. OLS	1. In-person=1 online=0	36.784***	58.306***	87.657***	93.068***
		(8.506)	(14.95)	(24.2)	(28.041)
	2. In-person=1 hybrid=0	21.288**	22.116	46.931*	41.207
		(10.556)	(21.318)	(27.259)	(31.323)
	3. Hybrid=1 online=0	15.496*	36.19**	40.726**	51.861**
		(8.252)	(16.567)	(15.953)	(22.382)
2. Mult. OLS	1. In-person=1 online=0	36.858***	37.161*	56.757**	24.678
		(13.083)	(19.142)	(28.883)	(36.541)
	2. In-person=1 hybrid=0	23.91*	16.166	31.045	9.016
		(12.831)	(21.865)	(27.445)	(32.789)
	3. Hybrid=1 online=0	3.926	8.976	12.243	15.946
		(9.359)	(13.767)	(13.597)	(27.215)
3. PSM	1. In-person=1 online=0	33.354***	24.244	43.527	-9.35
		(11.762)	(18.444)	(29.46)	(60.772)
	2. In-person=1 hybrid=0	27.735*	19.49	28.16	5.321
		(15.514)	(21.877)	(31.549)	(41.575)
	3. Hybrid=1 online=0	10.449	17.531	19.615	32.863
		(12.253)	(19.448)	(18.689)	(32.435)
4. NN	1. In-person=1 online=0	29.279**	23.173*	43.4*	57.066*
		(12.003)	(13.156)	(22.465)	(31.346)
	2. In-person=1 hybrid=0	22.998*	19.07	40.196	22.824
		(12.23)	(20.26)	(30.132)	(40.986)
	3. Hybrid=1 online=0	4.228	7.968	7.408	12.306
		(11.026)	(13.628)	(12.638)	(22.407)
5. KMDM	1. In-person=1 online=0	35.804***	46.616***	76.103***	80.343**
		(10.484)	(14.468)	(26.419)	(35.549)
	2. In-person=1 hybrid=0	22.926*	15.932	44.634	36.798
		(12.259)	(21.189)	(29.102)	(38.67)
	3. Hybrid=1 online=0	11.984	26.817*	25.913*	36.217
		(9.822)	(15.896)	(14.336)	(25.807)

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls for more than one college in a county were used by not reported. Biv.=Bivariate, OLS=Ordinary Least Squares, Mult.=Multiple, PSM=Propensity Score Matching, NN=Nearest Neighbor, KMDM=Kernel Multivariate Distance Matching.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 4. Average treatment effects on COVID-19 deaths

(Dependent variable: % change in deaths compared to deaths in two weeks before reopening date)

		(1)	(2)	(3)	(4)
Model	Y variable	0-15 Days	15-30 Days	30-45 Days	45-60 Days
1. Biv. OLS	1. In-person=1 online=0	8.328	20.179**	38.767***	47.623***
		(7.284)	(9.151)	(11.279)	(16.7)
	2. In-person=1 hybrid=0	11.968	24.573**	26.049**	31.625*
		(8.354)	(10.786)	(12.816)	(17.887)
	3. Hybrid=1 online=0	-3.639	-4.394	12.718	15.998
		(6.636)	(8.734)	(9.765)	(12.896)
2. Mult. OLS	1. In-person=1 online=0	3.975	11.827	21.892	23.073
		(11.306)	(14.333)	(15.466)	(23.729)
	2. In-person=1 hybrid=0	21.275**	21.38	18.535	33.799
		(10.503)	(14.31)	(14.962)	(21.19)
	3. Hybrid=1 online=0	-6.581	-.669	6.383	-9.539
		(8.087)	(10.439)	(11.451)	(14.376)
3. PSM	1. In-person=1 online=0	7.145	17.779	22.205	20.743
		(10.605)	(14.319)	(16.189)	(22.832)
	2. In-person=1 hybrid=0	24.792*	6.472	17.24	24.224
		(13.829)	(17.36)	(16.556)	(22.72)
	3. Hybrid=1 online=0	-9.36	-2.508	2.843	-12.06
		(8.408)	(10.887)	(12.536)	(16.792)
4. NN	1. In-person=1 online=0	12.327	8.927	2.846	-2.206
		(10.021)	(12.061)	(14.09)	(24.319)
	2. In-person=1 hybrid=0	13.62	2.361	12.037	22.998
		(11.963)	(18.819)	(15.225)	(18.622)
	3. Hybrid=1 online=0	-8.46	2.521	-2.041	-10.883
		(8.354)	(12.537)	(11.866)	(14.9)
5. KMDM	1. In-person=1 online=0	15.253*	26.339**	33.944***	38.431*
		(9.05)	(11.318)	(12.922)	(21.175)
	2. In-person=1 hybrid=0	15.947	18.682	19.245	35.242*
		(10.184)	(14.156)	(14.125)	(19.719)
	3. Hybrid=1 online=0	.618	7.894	15.029	1.367
		(7.834)	(10.975)	(11.398)	(15.997)

Note: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. Controls for more than one college in a county were used by not reported. Biv.=Bivariate, OLS=Ordinary Least Squares, Mult.=Multiple, PSM=Propensity Score Matching, NN=Nearest Neighbor, KMDM=Kernel Multivariate Distance Matching.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Table 5. Logistic regression coefficients for pairwise comparison.

Variables	(1) In-person=1 online=0	(2) Hybrid=1 online=0	(3) In-person=1 hybrid=0
% stayed home	-0.0405	-0.0589	0.0225
	(0.0383)	(0.0381)	(0.0329)
Log enrollment	-0.409***	-0.0650	-0.314**
	(0.149)	(0.155)	(0.137)
Log cost of attendance	0.451	0.336	0.0146
	(0.571)	(0.551)	(0.559)
Log endowment per student	-0.280***	-0.263**	-0.0662
	(0.107)	(0.117)	(0.0902)
Log student-faculty ratio	-0.459	-0.874**	0.318
	(0.335)	(0.396)	(0.323)
Public=1, 0 otherwise	-1.079*	-0.856	-0.386
	(0.556)	(0.541)	(0.535)
Log total population	-0.498***	-0.410**	-0.0859
	(0.188)	(0.188)	(0.167)
Black population (%)	0.00756	-0.00238	0.0197*
	(0.0111)	(0.0105)	(0.0115)
Log household income	0.290	0.247	0.215
	(0.563)	(0.540)	(0.572)
Republican votes (%)	0.0502***	0.0158	0.0323***
	(0.0131)	(0.0124)	(0.0119)
Mask ordinance=1, 0 otherwise	-0.690**	-0.368	-0.456*
	(0.315)	(0.341)	(0.260)
Log cases in Spring 2020	0.331*	0.301	0.0271
	(0.181)	(0.194)	(0.154)
Log deaths in Spring 2020	-0.171	-0.112	-0.0819
	(0.149)	(0.155)	(0.122)
Temperature in reopening month	-0.0253	-0.00453	-0.0286
	(0.0250)	(0.0217)	(0.0230)
Constant	5.165	4.480	1.073
	(8.633)	(8.786)	(8.108)
Observations (colleges)	532	456	538
Pseudo R-squared	0.231	0.075	0.082

Note: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1. The third category is omitted from the sample for each pairwise comparison. We use log forms of predictors that have few points larger than the rest of the data to reduce the influence of extreme values on coefficient estimates.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

5 Concluding remarks and implications

Our findings have several implications for higher education and health policy. First, results from logistic regression suggest that colleges with smaller endowments per student are less likely to choose online over in-person teaching mode, after controlling for other factors. Moreover, the average treatment effects show adopting an in-person mode instead of an online mode may increase COVID-19 cases. Therefore, colleges with small endowments need special policy attention to combat a disease-induced crisis.

Second, the analysis also suggests that campus reopening has a positive impact on COVID-19 cases across all teaching modes, and both new cases and new deaths tend to increase for colleges teaching in person as opposed to colleges teaching online. Since the treatment effects find counterfactuals, one can point out that the risk of spreading a communicable disease can be partially mitigated with an initiative from the colleges by increasing distance education elements in classes.

Finally, counties with more republican votes in Presidential Election 2016 were less likely to choose online or hybrid modes over in-person, ceteris paribus. The finding indicates that political affiliation might play a role in the containment of COVID-19 or any other rapidly communicable diseases.

Our study offers critical insights about choosing a teaching modality amid a disease-induced pandemic. Future research can look into the short-run and long-run effects of different teaching modalities on student learning outcomes.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

References

1. Abadie, Alberto, and Guido W. Imbens. 2011. "Bias-corrected matching estimators for average treatment effects." Journal of Business & Economic Statistics 29, no. 1 (2011): 1-11.
2. American Community Survey (ACS). 2020. "American Community Survey Data." Available https://www.census.gov/programs-surveys/acs/data.html, Accessed 10/01/2020.
3. Andersen, Martin S, Ana I Bento, Anirban Basu, Chris Marsicano, and Kosali Simon. 2020. "College Openings, Mobility, and the Incidence of COVID-19 Cases." medRxiv. (September 23: doi:10.1101/2020.09.22.20196048).
4. Atkeson, Andrew. 2020. "What will be the economic impact of COVID-19 in the us? Rough estimates of disease scenarios." National Bureau of Economic Research. No. w26867.
5. Austin, Peter C. 2011. "An introduction to propensity score methods for reducing the effects of confounding in observational studies." Multivariate Behavioral Research 46.3: 399-424.
6. Barnow, Burt S, GS Cain, and Arthur S Goldberger. 1981. "Issues in the Analysis of Selectivity Bias", Evaluation Studies Review Annal" (W. Stromdorfer and G. Farkas, Eds). Pg. 43-59. Sage Press, Beverly Hills.
7. Bureau of Transportation Statistics (BTS). 2020. "Daily travel during the COVID-19 public health emergency." Available https://www.bts.gov/browse-statistical-products-and-data/trips-distance/daily-travel-during-covid-19- pandemic, Accessed 10/01/2020.
8. Cohen, Jamie A, Dina Mistry, Cliff C Kerr, and Daniel J Klein. 2020. "Schools are not islands: Balancing COVID-19 risk and educational benefits using structural and temporal countermeasures." medRxiv. doi:10.1101/2020.09.08.20190942.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

9. Di Domenico, Laura, Giulia Pullano, G Pullano, Niel Hens, and Vittoria Colizza. 2020. "Expected impact of school closure and telework to mitigate COVID-19 epidemic in France." Technical Report. Epicx Lab.
10. Gallagher, Sean, and Jason Palmer. 2020. "The pandemic pushed universities online: the change was long overdue." Harvard Business Review, September 29. Available https://hbr.org/2020/09/the-pandemic-pushed-universities-online-the-change-was-long-overdue, Accessed 10/01/2020.
11. Gandolfi, Alberto. 2020. "Planning of school teaching during COVID-19." Physica D: Nonlinear Phenomena, 415(January 2021): 132753.
12. HealthData.gov. 2020. "COVID-19 state and county policy orders." https://healthdata.gov/dataset/covid-19-state-and-county-policy-orders, Accessed: 2020-09-30.
13. Huber, Martin, Michael Lechner, and Andreas Steinmayr. 2015. "Radius matching on the propensity score with bias adjustment: tuning parameters and finite sample behaviour." Empirical Economics 49, no. 1 (2015): 1-31.
14. Hubler, Shawn, and Anemona Hartocollis. 2020. "How colleges be- came the new COVID hot stops." The New York Times, September 11. Available https://www.nytimes.com/2020/09/11/us/college-campus-outbreak-covid.html, Accessed 10/01/2020.
15. Imbens, Guido W. 2000. "The role of the propensity score in estimating dose- response functions." Biometrika, 87(3): 706–710.
16. Jann, Ben. 2017. Kernel matching with automatic bandwidth selection. London, UK: Stata. Available https://www.stata.com/meeting/uk17/slides/uk17_Jann.pdf.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

17. King, Gary, and Richard Nielsen. 2019. "Why propensity scores should not be used for matching." Political Analysis, 27(4): 435-454.

18. Livadiotis, G. 2020. "Statistical analysis of the impact of environmental temperature on the exponential growth rate of cases infected by COVID-19". PLoS one, 15(5):e0233875.

19. McCaffrey, Daniel F, Beth Ann Griffin, Daniel Almirall, Mary Ellen Slaughter, Rajeev Ramchand, and Lane F Burgette. 2013. "A tutorial on propensity score estimation for multiple treatments using generalized boosted models." Statistics in medicine, 32(19): 3388–3414.

20. McKibbin, Warwick, and Roshen Fernando. 2020. "The economic impact of COVID-19." Economics in the Time of COVID-19, (Richard Baldwin and Weder di Mauro eds.), p. 45. London: CEPR Press.

21. MIT Election Data and Science Lab. 2018. County Presidential Election Returns 2000-2016 V6, Harvard Dataverse. 10.7910/DVN/VOQCHQ.

22. National Center for Education Statistics (NCES). 2020. "Integrated Postsecondary Education Data System." U.S. Department of Education. Institute of Education Sciences. Available https://nces.ed.gov/ipeds/use-the-data, Accessed 10/01/2020.

23. New York Times (NYTimes). 2020. "Coronavirus (COVID-19) Data in the United States." https://www.nytimes.com/interactive/2020/us/coronavirus-us-cases.html, Accessed: 2020-09-30.

24. National Centers for Environmental Information (NOAA). 2021. "County average temperature". Climate at a Glance: County Mapping. April. Available https://www.ncdc.noaa.gov/cag/county/mapping/50/tavg/202009/2/value, Accessed: 2021-04-01.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

25. Panovska-Griffiths, Jasmina, Cliff C Kerr, Robyn M Stuart, Dina Mistry, Daniel J Klein, Russell M Viner, and Chris Bonell. 2020. "Determining the optimal strategy for reopening schools, the impact of test and trace interventions, and the risk of occurrence of a second COVID-19 epidemic wave in the UK: a modelling study." The Lancet Child & Adolescent Health, 4(11): 817–827.
26. Picault, Julien. 2021. "Structure, flexibility and consistency: A dynamic learning approach for an online asynchronous course." Applied Economics Teaching Resources, 3(4): 30-43.
27. Rosenbaum, Paul R, and Donald B Rubin. 1983. "The central role of the propensity score in observational studies for causal effects." Biometrika, 70(1): 41–55.
28. Rosenbaum, Paul R, and Donald B Rubin. 1984. "Reducing bias in observational studies using subclassification on the propensity score." Journal of the American statistical Association, 79(387): 516–524.
29. Tyson, Alec. 2020. "Republicans remain far less likely than Democrats to view COVID-19 as a major threat to public health." Fact Tank by Pew Research Center, 1(1): 1. Available https://www.pewresearch.org/fact- tank/2020/07/22/republicans-remain-far-less-likely-than-democrats-to-view- covid-19-as-a-major-threat-to-public-health/, Accessed 10/24/2020.
30. Walke, Henry T, Margaret A Honein, and Robert R Redfield. 2020. "Preventing and responding to COVID-19 on college campuses." JAMA, September 29. doi:10.1001/jama.2020.20027.
31. Wrighton, Mark S, and Steven J Lawrence. 2020. "Reopening Colleges and Universities During the COVID-19 Pandemic." Annals of Internal Medicine. October. https://doi.org/10.7326/M20-4752.

medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .

Appendix

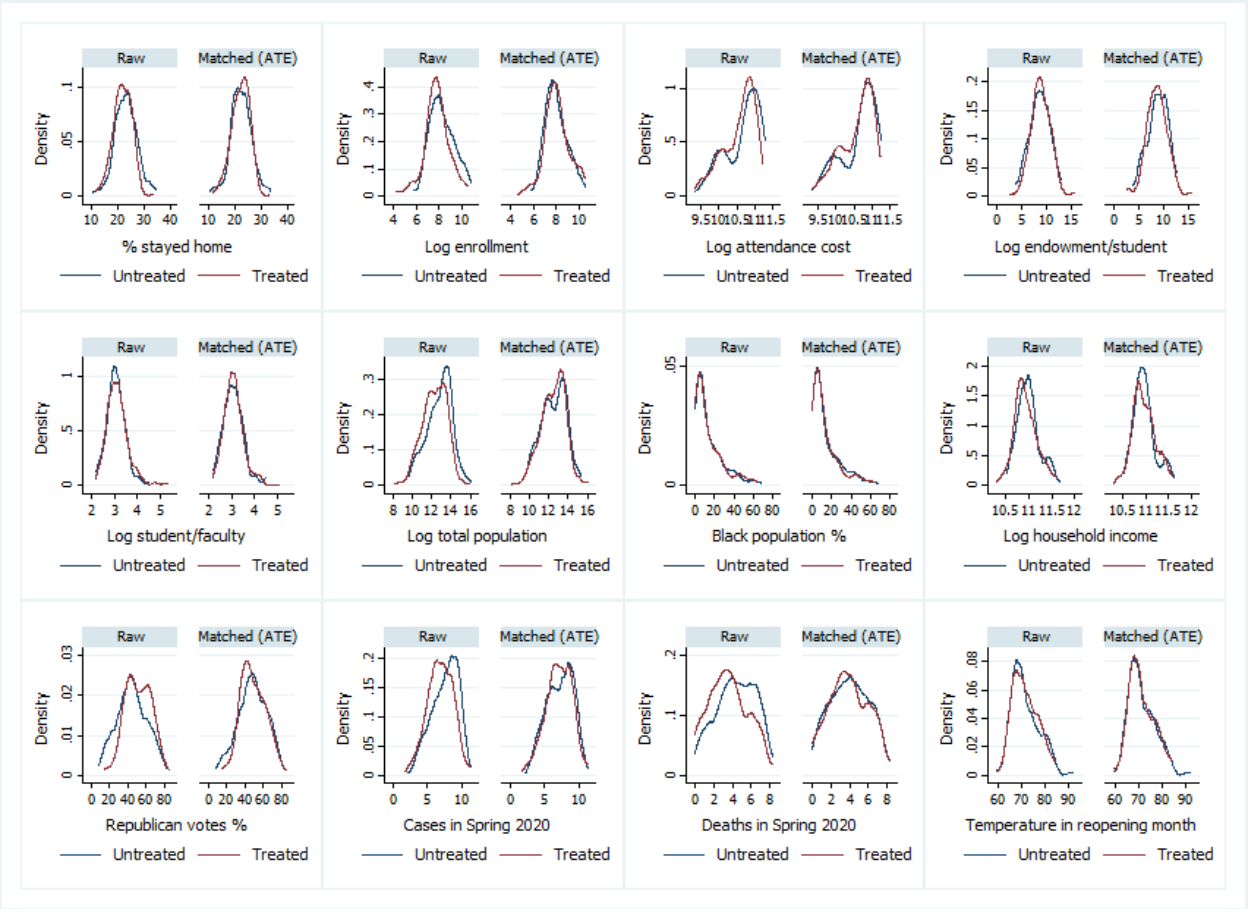
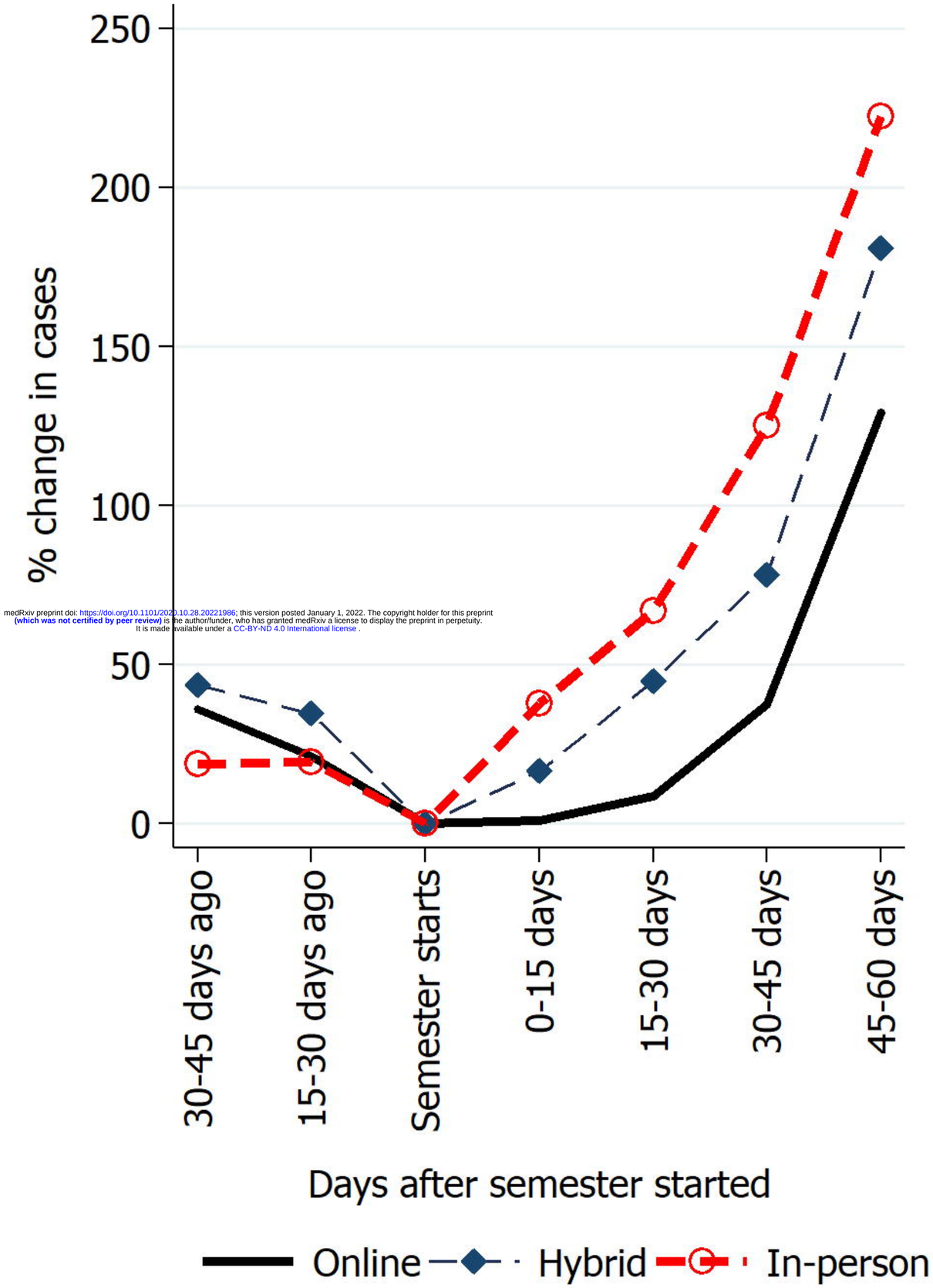
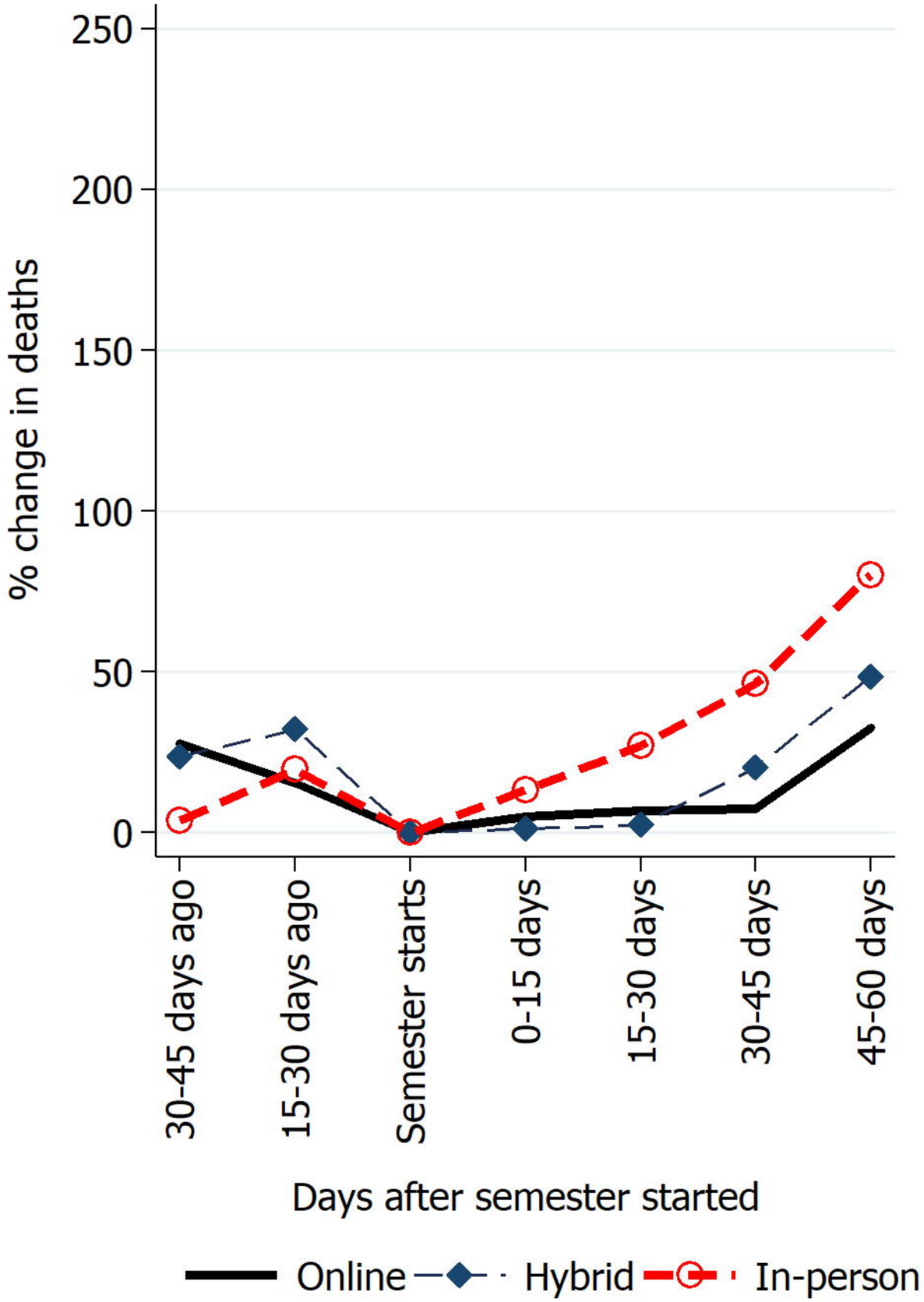


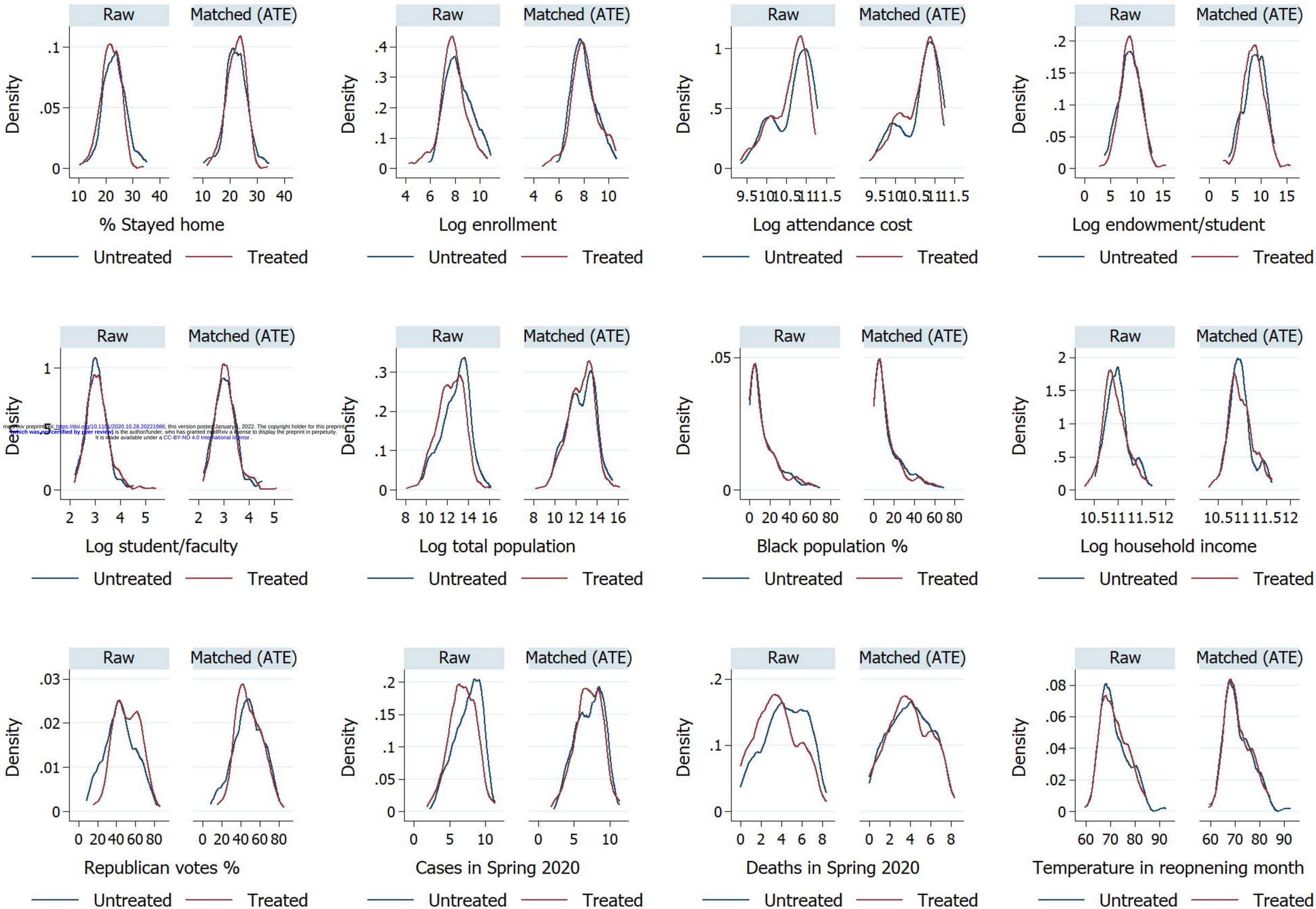
Fig. A1. Densities of major variables before and after matching.

Note: Figure shows kernel matching for multiple variables. See Table 1 for variable description and Table 2 for summary statistics.

medRxiv preprint doi: https://doi.org/10.1101/2022.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint
(which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity.
It is made available under a CC-BY-ND 4.0 International license .







medRxiv preprint doi: https://doi.org/10.1101/2020.10.28.20221986; this version posted January 1, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder, who has granted medRxiv a license to display the preprint in perpetuity. It is made available under a CC-BY-ND 4.0 International license .