Research Article

David Welsch*

The Impact of Mask Usage on COVID-19 Deaths: Evidence from US Counties Using a Quasi-Experimental Approach

https://doi.org/10.1515/bejeap-2021-0157
Received May 4, 2021; accepted October 27, 2021

Abstract: I examine the relationship between mask usage and COVID-19 deaths at the county level. When examining this relationship, even the direction caused by the potential endogeneity bias is unclear. In one direction, characteristics that are known to correlate with a larger amount of potential COVID-19 deaths, such as an older population, may make people more likely to wear masks. This will cause a bias that makes mask usage look less effective than it truly is. In the other direction, areas with higher risk tolerances may have less mask usage, but may at the same time be engaging in other behavior that puts them at higher risk for contracting COVID-19. This will cause a bias that makes mask usage look more effective than it truly is. The identification approach exploits a large set of controls and employs percentage of vote for Donald Trump in the 2016 election as an instrumental variable for mask usage. The main finding is that a one percentage point increase in the number of individuals who say they often or frequently wear a mask when within six feet of people will reduce COVID-19 deaths in a county by 10.5%, or six deaths in the average county.

Keywords: masks, face coverings, COVID-19, Corona virus

1 Introduction

Laboratory experiments have shown that masks can reduce respiratory droplets, which are believed to be the main transmission mechanism for the Corona Virus (Bahl et al. 2020; Lindsley et al. 2020; Verman, Dhanak, and Frankenfield 2020).

***Corresponding author: David Welsch**, Professor of Economics, University of Wisconsin - Whitewater, 800 S. Main St., 53190, Whitewater, WI, USA, E-mail: welschd@uww.edu

Due to studies of this nature and other evidence, many health officials have recommended the wearing of masks (face coverings) to mitigate the spread of the Corona Virus and resulting deaths due to COVID-19; in addition several areas have implemented orders requiring mask usage in public. However, many individuals are still reluctant to wear masks and are skeptical that masks can reduce COVID-19 related deaths. Many refuse to wear masks even when businesses or municipalities require it. In addition, many police departments seem reluctant to enforce mask mandates, often due to their own personal skepticism.[1] While the above laboratory studies are useful, they may not tell us how real-world mask usage will affect COVID-19 deaths; things that are true in laboratories don't always translate to real world situations that have more variables at play. Importantly, because of these concerns, skeptics about mask usage may be less likely to be convinced by the results of these laboratory studies.

To examine the effect of mask usage under real world conditions, I utilize a county level dataset and employ several empirical techniques to examine whether a larger percentage of the population wearing masks can reduce deaths from COVID-19. This paper demonstrates that a correctly specified model, which accounts for omitted variable bias (endogeneity), produces a result showing that mask usage reduces deaths from COVID-19. The effect size from this appears to be large. In addition, mask usage appears to lead to a greater reduction in COVID-19 deaths in counties where there are a large number of COVID-19 deaths.

There are two main empirical complications when examining the structural relationship between mask usage and COVID-19 deaths at the county level. It is also unclear which direction these complications will bias the estimates. The first is that an area's prior experience with COVID-19 will influence both mask wearing and will likely be correlated with future COVID-19 deaths. Areas that have had more COVID-19 deaths in the past may be more likely to have more individuals willing to wear masks. If these places are also more "naturally prone" to experience COVID-19 deaths, and if this is not accounted for, it would make mask usage look less effective than it truly is. The second empirical complication is related to the area characteristics. Here there are two main concerns, each of which would bias the coefficient in opposite directions. First, some areas may be more "naturally" prone to a greater amount of COVID-19 deaths; in other words, certain areas may have characteristics that make them more likely to have COVID-19 deaths. One example would be counties that have an older population, although

1 For example, anecdotal evidence on this see these newspaper articles: https://www.usatoday.com/story/news/politics/2020/09/16/covid-19-face-mask-mandates-go-unenforced-police-under-pressure/5714736002/, https://thehill.com/changing-america/opinion/509859-why-law-enforcement-isnt-enforcing-mask-mandates.

there are likely many other observable as well as unobservable characteristics that would also increase an area's risk factors. These counties may be areas where people are more likely to wear masks, but would likely have had more deaths (in the absence of this mask usage) than other similar areas that are less naturally prone to COVID-19 deaths. This again, if not accounted for, would make mask usage look less effective than it truly is. Second, some areas may have individuals with a higher overall tolerance for risk (less area risk aversion). This would likely make these areas less prone to mask usage but also more likely to have more COVID-19 deaths, because of other risky activities taken by these individuals. This would make mask usage look more effective than it truly is.

In order to address these complications I employ three main strategies. To account for prior COVID-19 experience, I include the number of COVID-19 deaths prior to the survey on mask usage. To account for area characteristics I employ two techniques. First, I am able to control for many observable characteristics, such as, but not limited to, the age profile and to some extent the historic health fragility of the population. Second, to account for any remaining unobserved characteristics that influence both mask usage and COVID-19 deaths, I implement the quasi-experimental approach of employing the percentage of individuals who voted for Donald Trump in the 2016 election as an instrumental variable (IV) for mask usage.

I find that a simple correlation shows a small but *positive* and statistically significant relationship between percentage of individuals in a county wearing masks and COVID-19 deaths. Once I control for county population and COVID-19 deaths prior to the survey this relationship shrinks but remains positive and statistically significant. Once a measure that captures riskiness and overall health in a county is included, this relationship becomes statistically insignificant. When my full set of county level controls are added, the relationship between mask usage and COVID-19 deaths becomes negative and statistically significant, but practically insignificant. In my preferred model that employs percentage of vote for Donald Trump in 2016 as an instrumental variable and a full set of control variables, I find that more individuals wearing masks more often reduces COVID-19 deaths. This result is both statistically significant and practically significant. Specifically, a one-percentage point increase in the number of individuals who say they often or frequently wear a mask when within six feet of other people will reduce COVID-19 deaths in a county by 10.5%, or about six people in an average county. I present evidence that the instrument is not weak and provide several arguments that the exclusion and monotonicity assumptions are satisfied. I show that the findings are robust across several different specifications. In addition, an examination of instrumental variable quantile regression results show that mask usage is more effective in counties with a large number of COVID-19 deaths. The main results are also robust to replacing COVID-19 deaths with COVID-19 cases.

2 Literature Review and Data

2.1 Literature Review

Not surprisingly, the existing research on the effectiveness of mask usage, while growing, is sparse. The few studies that do exist focus mainly on the ability of masks to filter aerosol particles or respiratory droplets (Bahl et al. 2020; Lindsley et al. 2020; Verman, Dhanak, and Frankenfield 2020). A limited number of econometric studies have examined the effect of mask usage on COVID-19 deaths or cases (Chernozhukov, Kasaha, and Schrimpf 2020; Karaivanov et al. 2020; Lyu and Wehby 2020; Mitze et al. 2020; Yilmazkuday 2020; Zhang et al. 2020). These studies generally find that mask usage reduces COVID-19 cases or deaths. Yilmazkuday has a similar dataset to this paper, however there are many important differences between the two papers, including his identification strategy; he relies on a difference-in-difference procedure in an attempt to identify the relationship between mask usage and COVID-19 deaths. There have also been some descriptive epidemiological analyses that attempt to forecast the results of different levels of mask usage. The most recent of these claims that 95% mask usage could mitigate the effects of a resurgence in many states (see IHME COVID-19 Forecasting Team 2020 among others).

Bundgaard et al. (2020) is the only randomized study, to the author's knowledge, that examines mask usage in the real world. They randomly assigned 3030 individuals in Denmark to the "treatment group" where they were encouraged to social distance, wear masks, and were supplied with 50 surgical masks, while 2994 were included as a control; this study took place early in the pandemic when few people were wearing masks. Their study examines whether social distancing and masks protect the wearer, but is unable to examine whether it protects others or slows the spread of the disease. While their results show that mask wearers were less likely to test positive for COVID-19, this result was not significant at the 5% level.

While there are now a growing number of studies that have demonstrated the effectiveness of mask usage on reducing the number of COVID-19 deaths, my paper makes an important contribution to the literature; my paper uses an unique identification strategy with a nationwide dataset. Employing different strategies to examine the same issue is important because replication of results is important, particularly in regards to mask usage where many individuals are still skeptical that they can help prevent COVID-19 deaths.

2.2 Data

The dataset in the preferred estimation consists of 3079 counties and county equivalents. This covers nearly every county in the U.S., excluding the small number that had a missing variable used in the analysis. As there are currently 3141 counties and county equivalents in the US, only 62 are omitted from the analysis.

The dependent variable of interest is deaths from COVID-19 at the county level. This was obtained from the nonpartisan initiative USAFacts which aggregated data from the CDC (Centers for Disease Control and Prevention) and state and local agencies.[2] I also use the USAFacts population estimates to measure county population. For COVID-19 deaths I use log of deaths (+1) per capita. I use log of deaths per capita in my preferred specification because it is much more "bell shaped" than any other transformation of this variable. However, I examine the effect of other transformations in the robustness check section.

Information on mask usage is from a survey conducted by Dynata at the request of the New York Times. It was a survey of approximately 250,000 interviews conducted from July 2 to July 14 2020; the methodology employed in the study ensures that mask usage is representative on the county level. Each individual was asked "*How often do you wear a mask in public when you expect to be within six feet of another person?*" Individuals were allowed to respond with: "never", "rarely", "sometimes", "frequently", or "always". I combine the percentage who responded with "frequently" or "always" into a single measure, which I use as the main variable of interest in the study. However, I also use other combinations of these in the robustness check section.

Information on the percentage of vote Donald Trump received in the 2016 election was obtained from the MIT Election Data and Science Lab. I use this percentage as an instrumental variable in the preferred specification.

Information on educational attainment, median income, minority demographics, and population density is from the US Census. Since information is not available for 2020, previous census information along with a combination of weighted averages and linear projections are used to predict the current value of many of these variables (more specific information on the construction of these variables is available from the author). The additional important control of overall mortality in 2016 is from the CDC mortality file.

The mean and standard deviation for all variables can be found in Table 1. The three main variables, COVID-19 deaths on September 1, mask usage, and

---

2 For information on how this variable was collected please see: https://usafacts.org/articles/detailed-methodology-COVID-19-data/.

Table 1: Descriptive statistics.

	Mean	SD
Dependent variables		
COVID deaths sept. 1	58.11	310.77
Per capita COVID deaths sept. 1 per 10,000	3.42	4.69
Log per cap. COVID deaths sept. 1	−8.22	1.02
Main variable of interest		
Mask usage frequently/always	71.58	13.11
Instrumental variable		
% Vote Trump	63.04	15.82
Control variables		
COVID deaths July 1	40.26	270.87
Per capita COVID deaths July 1 per 10,000	1.72	3.46
Log per cap. COVID deaths July 1	−8.22	1.02
Population in 10,000's	1.04	3.33
Log per cap. All deaths 2016	5.76	1.34
% College	21.66	9.53
% High school graduates	33.99	7.55
Percentage minorities	13.66	15.82
Percentage hispanic	9.47	13.73
Percentage female	49.90	2.24
Percentage age 20 to 29	12.31	3.17
Percentage age 30 to 39	11.70	1.66
Percentage age 40 to 49	11.62	1.41
Percentage age 50 to 59	13.86	1.52
Percentage age 60 over	25.78	5.65
Average household income	51,060.51	13,486.54
Population density (square miles)	242.61	1669.58

percentage of vote for Donald Trump, all exhibit a fair amount of variation. Note, that while September 1 may seem to be chosen arbitrarily, in the robustness check section I demonstrate that the results are not sensitive to this choice.

3 Model

The main empirical complication is omitted variable bias (endogeneity). First, counties with a larger amount of COVID-19 deaths in the past may have higher mask usage but at the same time be prone to more current COVID-19. This higher death rate could be due to factors that make this county more prone to COVID-19 or because a large amount of COVID-19 exists in the county. Second, there are other factors that affect both wearing a mask and COVID-19 deaths in a county. In one direction, counties with a “naturally” higher risk for COVID-19 deaths may both have more individuals wearing masks and have more COVID-19 deaths. This would cause mask usage to look less effective than it truly is. On the other hand, counties with a higher average level of risk aversion will have higher mask usage, but would likely have had less COVID-19 deaths due to other steps individuals would have taken to avoid exposure to COVID-19. This would cause mask usage to look less effective than it truly is.

The empirical approach attempts to correct for the above concerns in three ways. First, I utilize the timing of the survey on mask usage and when COVID-19 deaths are measured; I use overall COVID-19 deaths up until approximately six weeks after the survey is completed as the outcome variable, while controlling for the number of deaths prior to the survey. Second, I incorporate a rich set of county level controls. Third, I employ an external “instrument” that, once other factors are controlled for, correlates with mask usage but not with any remaining unobservables that are correlated with COVID-19 deaths.

The remainder of this section presents the empirical approach. The first subsection outlines the baseline equation, followed by discussion of my identification approach.

3.1 Baseline Model

The baseline model attempts to address endogeneity through two of the channels mentioned above, timing of the survey and timing of measured COVID-19 deaths, along with a large set of county controls. The formal version of the baseline model takes the following form:

$$\begin{aligned}\mathrm{COVIDdeaths}_{\mathrm{Sep1}} = \delta_0 + \delta_1 \mathrm{Mask}_{\mathrm{July}} + \delta_2 \mathrm{COVIDdeaths}_{\mathrm{July1}} \\ + \delta_3 \mathrm{Mort}_{2016} + \boldsymbol{x}'\boldsymbol{\beta} + \mu\end{aligned}$$

where $\mathrm{COVIDdeaths}_{\mathrm{Sep1}}$ is the log of number of COVID-19 deaths +1 in a county per capita up until and including September 1, $\mathrm{Mask}_{\mathrm{July}}$ is the percentage of individuals in a county that responded that they frequently or always wear a mask in public when they expect to be within six feet of another person, $\mathrm{COVIDdeaths}_{\mathrm{July1}}$is the log of number of COVID-19 deaths +1 in a county per capita up until and including July 1, and Mort_{2016} is the log of overall total deaths in a county +1 per capita in 2016.[3] Additional controls are in the vector $\boldsymbol{x}$; which includes population of

3 Any choice of date for the dependent variable would necessarily be arbitrary. I choose September 1st because it is the first of a month and not too far removed from the survey; this is an attempt

the county, educational attainment in the county, percentage of minorities in a county, percentage Hispanic in a county, percentage female in a county, the age demographics of a county, average household income of a county, and population density; the specifics of these can be found in Table 1. The following are estimable coefficients: $\delta_0, \delta_1, \delta_2, \delta_3, \beta$, while μ is the error term.

By including COVID-19 deaths prior to the survey on mask usage, the endogeneity concern above is mitigated to some extent. This allows me to control how much the county has been affected by COVID-19 in the past. It would be expected that a county with a higher amount of deaths prior to the survey would have more mask usage in the future and at the same time a larger amount of COVID-19 deaths up until and including September 1st.

The other controls also help to alleviate the endogeneity concern. While some of the controls, such as age categories and percentage minority, are included for obvious reason, some do bear discussing. Log of all deaths per capita in 2016 is included because it accounts for both the "historic" amount of risk aversion and the fragility of the population. Education variables and median income are included because more educated or higher income individuals may be more or less likely to wear masks and may at the same time be more or less likely/able to engage in other COVID-19 prevention measures. Robust standard errors are used in all estimations.

Even with the rich set of controls and the timing of the survey there may still be some endogeneity concerns. If there are any remaining unobservables that affect both mask usage and deaths due to COVID-19, or the controls do not fully account for the risk aversion of the population or the natural level of COVID-19 risk in a county, estimates of δ_1 will still be biased. In the subsequent subsection I use a quasi-experimental method in an attempt to fully identify the model.

3.2 Instrumental Variable Approach

The identification approach relies upon instrumental variable techniques where Eq. (1) above becomes the second stage and the following is the first stage equation:

$$\mathrm{Mask}_{\mathrm{July}} = \alpha_0 + \alpha_1 \mathrm{CovidDeaths}_{\mathrm{July1}} + \alpha_2 \mathrm{Trump16} + \alpha_3 \mathrm{Mort}_{2016} + \boldsymbol{x}'\boldsymbol{\gamma} + \varepsilon \tag{2}$$

where Trump16 is the percentage of individuals in the county that voted for Donald Trump in the 2016 election, $\alpha_0, \alpha_1, \alpha_{2,}, \alpha_3$, $\boldsymbol{\gamma}$ are estimable coefficients, and ε is the error term. The other variables are defined above.

to make sure that attitudes in the counties may not have changed much. In the robustness check section I demonstrate that the choice of date does not affect the results.

Trump16 is the excluded instrumental variable. Instrument validity rests on three assumptions. First Trump16 must significantly correlate with mask usage conditional on other control variables, and this must be powerful (instrument strength). There is strong documentation that some President Trump supporters may be influenced by his comments to not wear masks as is illustrated by the following quotes: "The C.D.C. is advising the use of nonmedical cloth face covering as an additional voluntary public health measure. So it is voluntary. You don't have to do it. They suggested for a period of time, but this is voluntary. I don't think I'm going to be doing it". (April 3, at the White House) and "I don't agree with the statement that if everybody wears a mask, everything disappears" (July 19, to Fox News host Chris Wallace).[4]

Instrument strength is also testable. Staiger and Stock (1997) and Stock and Yogo (2005) set the often cited benchmark for an instrument's strength. They argue that the first stage F-statistic of the excluded instrument must be, at a minimum, 10 or 16.38 respectively. To test this condition, I examine the estimation from Eq. (2). These results appear in the Appendix Table A1. The F-stat for the excluded instrument is approximately 435.07 and 49.73 in the models without and with the full set of controls respectively, thus exceeding the threshold conventional benchmark for power. In a recent paper, Lee et al. 2020 demonstrate that the "F threshold" either should be increased to 104.7 or if kept at 10, the critical value for 5% significance needs to be increased to 3.43 for the second stage endogenous variable's coefficient. In the preferred model the F-stat does not reach this 104.7 threshold, however the t-stat in the main result, found later, is 6.26, exceeding this new 3.34 critical value.[5]

The second assumption holds that the instrument must be conditionally uncorrelated with the error term in Eq. (1) (exclusion restriction), or more loosely with the amount of COVID-19 deaths on September 1. This assumption cannot be tested directly in my exactly-identified models so I rely on the instrument's intuitive appeal for validity. Recall the two potential omitted factors of concern are the risk tolerance of the population and the susceptibility of the county's population to contracting and dying from COVID-19. In other words, the percentage of individuals who voted for Donald Trump in 2016 should not be *directly* correlated with amount of risk aversion in a county or the "natural tendency" for individuals to acquire and die from COVID-19 *once* other factors are controlled for. With

4 These quotes can be found in the New York Times Article "*In His Own Words, Trump on the Coronavirus and Masks*" by Daniel Victor, Lew Serviss and Azi Paybarah on October 2, 2020.

5 More technically, they present formulas on how to use the value of the F-statistic to adjust the critical values. Using my first stage F-statistic value of 49.73, the new 5% critical value is 2.16, which again is smaller than my t-stat.

regard to both, but more so the former, I note that I control for a county's previous (overall) death rate. In addition, it seems unlikely, especially after controlling for things such as the age of the population of a county, that places with different percentages of votes for Donald Trump are more or less likely to have a natural propensity for COVID-19 deaths or a different level of risk aversion. In Section 5, I also perform some tests to investigate the relationship between percentage of vote for Donald Trump and historic levels of death rates due to injury.

The third assumption required is monotonicity. For monotonicity to hold, the percentage of vote for Donald Trump is allowed not to impact mask usage in some counties, but in all counties where it has an impact this effect must be negative ($\alpha_2 \leq 0$); in other words, there are no "defiers". Table A2 show that the estimated coefficient $\hat{\alpha}_2$ is negative, and while this does not show it would be true for the entire population, it seems unlikely that, everything else equal, if a specific county spontaneously had more support for Donald Trump in 2016 it would increase mask usage. It does seem plausible that in some counties, such as counties with many COVID-19 cases, if support for Donald Trump were to be increased it may not reduce mask usage (having "never-takers"), however this does not violate the monotonicity assumption. If there are some counties where the percentage of vote for Donald Trump does not impact mask usage, this would mean the result found is an LATE (local average treatment effect). Having identified an LATE simply means the relationship revealed would not apply to counties where changing the amount of support for Donald Trump does not affect mask usage. This implies that the results found later *may* not hold in counties with a large amount of mask usage and COVID-19 cases, where most people wear a mask and mask usage is not influenced by political affiliation.

Note that the instrument I utilize may not be able to circumvent the issue that mask wearing may be correlated with other COVID-19 prevention methods and regional strategies to fight COVID-19, such as participating in less large social gatherings. *If* these other preventive measures are negatively correlated with the percentage of individuals who voted for Donald Trump, the effect of mask usage on COVID-19 deaths will be overstated. However, even if the exclusion restriction is partially violated in this way, it would only result in some attenuation bias, it will likely not fully negate the results. In addition, I investigate this possibility in Section 5 by employing a crude measure of social distancing and find this would not qualitatively alter the results.

I also perform the Hausman-WU-Durbin test for endogeneity, which is robust to heteroskedasticity. It is important to note that this test relies on the fact that one has identified a valid instrument. The null hypothesis of this test is that the variable of interest is not endogenous. The p-value from this test can be found at the bottom of Table 2. It shows, given that the instrument is valid, that there is a strong indication of endogeneity.

4 Results

In addition to the main IV model I also estimate three OLS models with various controls and an IV model with fewer controls. A full set of results are included in the Appendix Table A1 and a summary of the main results can be found in Table 2. This section begins with a brief discussion of estimates of the control variable's coefficients before moving to the main point of interest: the influence of mask usage on COVID-19 deaths.

4.1 Control Variables

Most control variables have the associations one would expect from basic correlations and anecdotal evidence. The full results for the five main estimations can be found in the Appendix Table A2. In this subsection, I focus on the results found in columns (4) and (6) which include a full set of controls for OLS and IV estimates respectively. In OLS estimates, counties with a larger percentage of college graduates (relative to the percentage of individuals with no high school degree) have fewer deaths due to COVID-19. Counties with a larger percentage of minority individuals, Hispanic individuals, and women all have a larger percentage of deaths due to COVID-19. A larger percentage of individuals over the age of 60, between 20–29, and 40–49 (all relative to the percentage of individuals that are less than 20 years old) are all associated with a larger percentage of deaths due to COVID-19. Interestingly, in the OLS estimates a larger percentage of individuals 50–59 and a larger population density is associated with fewer COVID-19 deaths.

Additionally, the IV estimations with all the controls emphasize the importance of more fully accounting for endogeneity. The counterintuitive sign on the estimated coefficient for percentage of individuals 50–59 found above now has the expected positive sign and is statistically significant under IV, and several other coefficients such as percentage of individuals over the age of 60 are much larger in magnitude. This may demonstrate the problem of so-called "smearing", the effect that when the endogeneity of one variable is not accounted for, all coefficients are inconsistent, not just the one associated with the endogenous variable.

Table 2: Main results (dependent variable: log per capita COVID deaths on and before September 1).

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)	IV (6)
Mask usage frequently/always	0.007***	0.004***	0.001	−0.004***	−0.009***	−0.106***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.017)
Log per cap. COVID deaths July 1		0.724***	0.742***	0.697***	0.754***	0.739***
		(0.010)	(0.011)	(0.010)	(0.012)	(0.022)
Population		✓	✓	✓	✓	✓
Log all deaths per capita 2016			✓	✓	✓	✓
Other controls				✓		✓
F-test of exclud. instr. in 1st stage					435.07	49.73
Hausman-WU-Durbin F-stat p-value					0.0005	0.0000

All estimations also include a constant. The estimated coefficients for other controls can be found in Table A2. Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

4.2 Mask Usage and COVID-19 Deaths

I turn now to examining the effect a larger percentage of a county's population wearing masks has on deaths from COVID-19. I have two main objectives: first, to demonstrate that incorrect modeling leads to incorrect results and second, to examine both the direction and size of effect that mask usage has on COVID-19 related deaths. Table 2 summarizes these main results.

The simple OLS estimation that includes no controls reveals a positive and significant relationship between mask usage and COVID-19 deaths (result (1)). When COVID-19 deaths on July 1 and population are added as controls. this relationship shrinks by nearly half but remains positive and statistically significant (result (2)). Furthermore when total deaths in a county in 2016 is added (to some extent accounting for the overall risk aversion and fragility of the population), the coefficient on mask use loses both size and statistical significance, but remains positive (result (3)). In the fourth OLS estimation that includes the full set of controls, the coefficient on mask usage becomes negative and highly statistically significant, but is small in magnitude (result (4)). In the final two estimations presented in this table I further account for endogeneity through IV. When IV is used with limit controls the coefficient on mask usage remains negative and statistically significant and more than doubles in magnitude (result (5)), relative to OLS with a full set of controls.

Finally, in the preferred specification, when both a full set of controls are included and the IV strategy is implemented, the coefficient on mask usage is negative, statistically significant, and large in magnitude (result (6)). In particular, in result (6) the coefficient is 21 times larger in magnitude than the coefficient in model (4), which is OLS with all of the control variables. This result indicates that implementing IV substantially alters the average effects of mask usage on deaths.

Specifically, the coefficient on mask usage in model (6) indicates that a one percentage point increase in the number of individuals in a county who always or frequently wear masks when within six feet of others is associated with a 10.5% decrease in deaths related to COVID-19. To get some indication of effect size I examine this at mean values of all variables. If the value of the mask usage variable would increase from the mean of 72–73%, expected deaths decrease by approximately six (from 58 to 52).[6] This result appears to remain large even if I use the conservative estimate of taking the bottom of the 99% confidence interval. The bottom end of the 99% confidence interval for the coefficient on mask usage is approximately 0.06, which would translate into approximately three and a half

6 When taking the partial derivatives of COVID-19 deaths per capita and mask usage and solving for the coefficient, the population variable will cancel out.

lives saved in the average county, for a one percentage point increase in the mask usage variable.

5 Robustness Checks and a Further Examination of Excludability

5.1 Robustness Checks

The robustness checks show that the results are not sensitive to different functional forms, the date when COVID-19 deaths are measured and measures of mask use; in addition, I examine results when replacing the dependent variable of COVID-19 deaths with COVID-19 cases. The results for these robustness checks that alter functional form and measures of mask usage can be found in Table A3; all the results in this table are estimated using the methods of estimation (6) from Table 2, full controls and the IV estimation procedure, only the specification of the dependent variable is changed. The first three estimates found in this table change the way the dependent variable on COVID-19 deaths is measured; it is first measured as per capita without logging (now scaled at per 10,000 individuals) (result (1)), next as a level variable not measured per capita or logged (result (2)), and as log variable not in per capita terms (result (3)). Note in these three estimations the two other measures of deaths, COVID-19 deaths on July 1 and all deaths in 2016, are measured in the same manner as the dependent variable. Next, I return to measuring the dependent variable in log per capita, but alter other aspects of the model. First, the dependent variable of deaths is now the amount of *new* deaths between September 1 and July 1 (result (4)), then the amount of deaths as of the most current date available (result (5)), then the difference between the amount of deaths currently available and amount of deaths on July 1 (result (6)).[7] Finally, I return to the original measure of the dependent variable and alter the mask usage variable; first I limit it to the percentage of individuals who say they "always wear a mask. . . " (result (7)) and then expand it to those who either say they "always, frequently, or sometimes wear a mask. . . " (result (8)).

None of the results are qualitatively or quantitatively much different from the main results. In fact if I choose the amount of new deaths between September 1 (or the current date) and July 1, it increases the size of the coefficient on the mask use variable. The result from model (2) may seem to slightly contradict the six death approximation found in Section 4, however this must be interpreted with

7 In this version of the paper the most current COVID-19 death number is from November 3.

caution since this specification of the dependent variable is not bell-shaped and it may not capture the mean county's results.

Table A4 examines estimations where the dependent variable of log per capita of COVID-19 deaths is replaced with log per capita of COVID-19 cases. These results should be interpreted with caution. The measure of cases depends heavily on testing availability in an area along with individuals' willingness to get tested. This is important because testing availability and willingness is likely correlated with an area's responsiveness to COVID-19 along with how impacted the area is by COVID-19. Importantly, an area that is more responsive to COVID-19 may have more "cases" (positive tests) than one that is less responsive because they have widespread testing; however, their true measure of COVID-19 cases may be lower because of other preventive measures they are employing. In addition, deaths is a more important outcome measure for several reasons, including that most individuals have mild forms of COVID.

These issues notwithstanding, the results for cases are qualitatively similar to the main results for deaths. In the same way as in the COVID-19 deaths estimations, the simple OLS regression that includes no controls shows there is a positive and significant relationship between mask usage and COVID-19 cases (result (1)). In the more complicated models this relationship becomes negative, statistically significant, and large. The main difference is that in the COVID-19 cases estimation the coefficient becomes negative and statistically significant "earlier" as you add to the models. In the main model (result (6)) the coefficient sizes are highly similar: 0.106 for the COVID-19 deaths estimation and 0.103 for the COVID-19 cases estimation.

5.2 Examination of the Excludability Assumption

As stated above, I am unable to formally test for instrument excludability: is the instrument correlated with remaining unobservables once control variables are considered? Of particular concern is the extent that a county's natural proclivity towards COVID-19 deaths and the risk tolerance are correlated with percentage of vote for President Trump. As stated above it is not possible to test this formally, however, I am informally able to examine the relationship between the instrument and a county's historic fragility/risk aversion by examining the relationship between my instrument and historic death rate. There is a positive but very weak association between percentage of vote for Donald Trump in 2016 and overall deaths in 2016. A one percentage point increase in the vote for Trump is associated with 0.0000676 more (overall) deaths in 2016. When county population and other factors are taken into account this number decreases to 0.000021 deaths. This association being weak and the fact that historic death rate is included as a control should alleviate much of the concern that the exclusion restriction is violated in this way. In addition, the association between percentage of vote for Donald Trump in 2016 and *injury* related deaths in 2016 has a negative association, possibly indicating that counties with more overall support for Donald Trump actually have more average risk aversion not less.

An additional concern is that other COVID-19 preventative measures are excluded from the analysis, such as amount of social distancing in a county. This is problematic if these factors are strongly correlated with mask usage *and* the percentage of vote for Donald Trump in 2016. To investigate this possibility I use a measure of social distancing, the difference of average distance traveled by individuals in a county on November 1 2020 relative to average distance traveled by individuals in that same county on the same day of the week in a non-COVID-19 time period (the larger this number the less relative social distancing).[8] I standardize this variable since it has a bell-shaped distribution, and doing so will give some indication of size in the interpretations. With this measure, I first examine the relationship between it and percentage vote for Donald Trump. While the relationship between these two variables is statistically significant, it appears to be small in magnitude: a one percentage point increase in the number of votes for Donald Trump is associated with an approximately 0.02 standard deviation increase in this relative distance measure (less relative social distancing). When other factors are controlled for this relationship is reduced to a 0.006 standard deviation increase.

Now I examine if including this measure of social distancing as an additional control in my main estimation changes the results. I do not use this variable as a control in my main results because this measure is missing for many counties and reduces the sample size by over 10%. These results can found in Table A5. The first important result is that, in the first stage the relationship between mask usage and social distancing is statistically insignificant. This demonstrates there is no evidence of a statistical relationship between this potentially omitted variable and the endogenous variable. Also note from the first stage, the F-statistic for the coefficient of percentage of vote for Donald Trump actually increases with this additional control; however, this F-statistic comparison should be interpreted with caution because of the change in sample size. Most importantly in the second stage, while the coefficient on mask usage does decrease by a small amount when the social distancing measure is included, the results are still qualitatively

8 This number is available from Unacast, and a description of this calculation can be found at: https://www.unacast.com/post/the-unacast-social-distancing-scoreboard.

the same. This indicates that the results are robust to this specification change. Finally, while the coefficient on social distancing in the second stage is statistically significant, it is only significant at the 10% level, indicating that this is potentially a weak predictor of COVID-19 deaths.

6 Further Investigation

To get a more complete picture of the relationship between mask usage and deaths, I examine the effect of mask usage for different levels of COVID-19 deaths. First, I examine how mask usage affects COVID-19 deaths differently depending on past levels of COVID-19 deaths. Second, I examine this effect at different levels of the distribution of "current" deaths.

To examine the effect mask usage has at different levels of previous deaths, I add the interaction term of the mask usage variable multiplied by the number of deaths (logged per capita) on July 1. In the case of this interaction model, I treat both mask usage and the interaction term as endogenous and use both percentage of vote for Donald Trump in 2016 and the interaction of this with log of COVID-19 deaths per capita on July 1 as excluded instruments. This is admittedly an ad hoc solution for finding the additional needed instrument.

Next, I use quantile regression techniques. Quantile regression gives information about the relationship between mask usage and COVID-19 deaths at different points of the conditional distribution of COVID-19 deaths. To estimate quantile regression with instrumental variable techniques, I implement the method suggested by Machado and Silva (2019). They devise a moment-based approach. Standard errors are calculated using the nonparametric bootstrap procedure with 200 repetitions.

Table 3 presents the results for the interaction term estimation, as well as (instrumental variable) quantile regressions for the 0.1, 0.2, … 0.9 quantiles. The interaction result indicates that if the previous death rate was larger, mask usage will have a greater ability to reduce deaths. The quantile regression results demonstrate that mask usage has a greater ability to reduce deaths in areas with a large number of COVID-19 deaths. Specifically, at the 0.9 quantile the effect of a one percentage point change in the number of individuals who frequently or always wear masks is 12.8%, while at the 0.1 quantile it is 8.5%.[9]

9 Note, the median regression result (0.5 quantile) yields nearly identical results to the mean regression result found earlier. This is not surprising because the distribution of the main outcome variable used is highly symmetric.

Table 3: Interaction and quantile estimations.

	Interaction	Quantile regressions								
		0.10	**0.20**	**0.30**	**0.40**	**0.50**	**0.60**	**0.70**	**0.80**	**0.90**
Mask usage	−0.164***	−0.085***	−0.091***	−0.095***	−0.100***	−0.104***	−0.108***	−0.114***	−0.119***	−0.128***
	(0.037)	(0.029)	(0.025)	(0.023)	(0.021)	(0.019)	(0.018)	(0.017)	(0.018)	(0.024)
Mask * deaths July1	−0.007**									
	(0.003)									

All estimations are from second stage IV regressions and include the full set of controls, which can be found in Table A2. The quantile regression standard errors are from the nonparametric bootstrap with 200 repetitions, and are found in the parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

7 Conclusions

There is some skepticism about the efficacy of mask usage on COVID-19 outcomes. This study uses county level data to examine the effect of mask usage on COVID-19 deaths. Examining this relationship is difficult because of the endogeneity or omitted variable bias that occurs. That is, there are factors that are correlated with both mask usage in a county and COVID-19 deaths. In one direction, low average risk aversion in a county will likely be associated with more COVID-19 deaths and may also be associated with fewer individuals wearing masks. In the other direction, counties with a higher "natural" risk of COVID-19 deaths may have more individuals on average willing to wear masks and also have more COVID-19 deaths. Thus, even the direction of this bias is ambiguous.

In an attempt to identify this relationship, I use a rich set of controls, including COVID-19 deaths prior to the survey on mask usage, and an instrumental variable technique that employs percentage of vote for Donald Trump as an instrument for mask usage. The main finding is that a one-percentage point increase in mask usage would decrease COVID-19 deaths by approximately 10.5% or approximately six people in the average county. I also find that mask usage is more effective at reducing deaths in areas where there are more deaths due to COVID-19.

One potential caveat of this study should be noted. It is possible that mask usage could be correlated with general "COVID-19 cautiousness" or lack thereof and regional strategies to fight COVID-19. It could be that places that have more mask usage also have more social distancing, less social gatherings, and more likely quarantine or isolate when ill, after a positive test, etc. If the instrument used is also strongly correlated with those activities as well, it would lead to my results overstating the effectiveness of masks on reducing deaths. I attempt to explore this possibility and find little evidence that this is the case. However, even if the exclusion restriction is partially violated in this way it will only lead to some attenuation bias, but will be unlikely to negate the main result showing the efficacy of mask usage. In addition, I show that including a crude measure of social distancing will not qualitatively alter my results. This potential caveat notwithstanding, future work should include a thorough cost-benefit analysis based on the results from this paper.

It has been difficult for policy makers to incentivize mask usage and greater research on how to do so is needed. However, my hope is that the results from this paper lead to greater belief in mask usage, both by policy makers and the public at large. In addition, I hope that researchers in other fields, such as psychology, spend greater time and effort on determining ways of encouraging or "nudging" individuals into more mask usage.

Acknowledgments: The author would like to thank Nikki Brendemuehl for research assistance. For helpful feedback and comments, the author thanks Benjamin Artz, Jason Baron, Nicholas Lovett, Matthew Winden, and David Zimmer.

Appendix: Additional Tables

Table A1: First stage results (mask usage frequently/always).

	1st Stage	1st Stage
% Vote Trump	-0.301^{***}	-0.130^{***}
	(0.014)	(0.018)
Log per cap. COVID deaths July 1	0.784^{***}	0.547^{***}
	(0.178)	(0.172)
Population per 100,000	-0.072^{*}	-0.264^{***}
	(0.037)	(0.048)
Log per cap. All deaths 2016	2.938^{***}	2.251^{***}
	(0.195)	(0.202)
% College[a]		0.032
		(0.045)
% High school graduates[a]		−0.038
		(0.044)
Percentage minorities		0.138^{***}
		(0.017)
Percentage hispanic		0.392^{***}
		(0.018)
Percentage female		0.590^{***}
		(0.119)
Percentage age 20 to 29[b]		1.802^{***}
		(0.125)
Percentage age 30 to 39[b]		0.282
		(0.215)
Percentage age 40 to 49[b]		2.722^{***}
		(0.206)
Percentage age 50 to 59[b]		1.566^{***}
		(0.203)
Percentage age 60 and over[b]		1.215^{***}
		(0.085)
Average household income (1000s)		0.213^{***}
		(0.025)
Population density (square miles)		$-3.34 \times 10^{-4***}$
		(7.59×10^{-5})
F-stat of the excluded instrument	435.07	49.73

All estimations also include a constant. Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. [a]The reference group is % with no high school degree. [b]The reference group is % under the age of 20.

Table A2: Full set of results.

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)	IV (6)
Mask usage frequently/always	0.007***	0.004***	0.001	−0.004***	−0.009***	−0.106***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.017)
Log per cap. COVID deaths July 1		0.724***	0.742***	0.697***	0.754***	0.739***
		(0.010)	(0.011)	(0.010)	(0.012)	(0.022)
Population per 100,000		0.009***	−0.006**	−0.009***	−0.005*	−0.034***
		(0.003)	(0.003)	(0.002)	(0.003)	(0.007)
Log per cap. All deaths 2016			0.078***	0.076***	0.119***	0.328***
			(0.009)	(0.010)	(0.017)	(0.048)
% College[a]				−0.012***		7.74×10^{-6}
				(0.002)		(0.005)
% High school graduates[a]				6.95×10^{-5}		−0.002
				(0.002)		(0.005)
Percentage minorities				0.012***		0.033***
				(0.001)		(0.004)
Percentage hispanic				0.017***		0.061***
				(0.001)		(0.008)
Percentage female				0.040***		0.102***
				(0.007)		(0.018)
Percentage age 20 to 29[b]				0.021***		0.218***
				(0.007)		(0.036)
Percentage age 30 to 39[b]				0.004		0.043
				(0.011)		(0.026)
Percentage age 40 to 49[b]				0.093***		0.349***
				(0.011)		(0.048)

Table A2: (continued)

	OLS (1)	OLS (2)	OLS (3)	OLS (4)	IV (5)	IV (6)
Percentage age 50 to 59[b]				-0.055^{***}		0.140^{***}
				(0.010)		(0.040)
Percentage age 60 and over[b]				0.026^{***}		0.154^{***}
				(0.004)		(0.023)
Average household income (1000s)				0.001		0.023^{***}
				(0.001)		(0.005)
Population density (square miles)				$-1.10 \times 10^{-5*}$		$-4.08 \times 10^{-5***}$
				(6.47×10^{-6})		(1.25×10^{-5})
Number of observations	3142	3142	3103	3096	3079	3079

All estimations also include a constant. Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. [a]The reference group is % with no high school degree. [b]The reference group is % under the age of 20.

Table A3: Robustness checks (altering functional form and measures of mask usage).

	Per cap Not logged (1)	Not per cap Not logged (2)	Not per cap Logged (3)	Log per cap Between 7–1 and 9–1 (4)	Log per cap Current (5)	Log per cap Between 7 and 1 and current (6)	Log per cap Mask always only (7)	Log per cap Mask always, freq., and sometimes (8)
Mask usage	−0.130***	−1.456*	−0.100***	−0.171***	−0.200***	−0.170***	−0.076***	−0.150***
	(0.037)	(0.744)	(0.016)	(0.026)	(0.047)	(0.025)	(0.010)	(0.026)
COVID deaths 7 – 1	1.081***	0.913***	0.750***	0.281***	0.610***	0.193***	0.733***	0.729***
	(0.023)	(0.022)	(0.022)	(0.034)	(0.059)	(0.032)	(0.018)	(0.022)
Pop in 100,000s	−0.017	29.138***	−0.033***	−0.009	−0.149***	−0.020**	−0.027***	−0.038***
	(0.011)	(6.255)	(0.007)	(0.010)	(0.038)	(0.009)	(0.005)	(0.009)
All deaths 2016	0.006	−0.003	0.564***	0.165**	2.265***	0.254***	0.312***	0.329***
	(0.004)	(0.008)	(0.043)	(0.074)	(0.139)	(0.071)	(0.039)	(0.051)
% College[a]	−0.002	−0.088	4.76×10^{-5}	0.005	−0.044***	−0.003	−0.003	−0.004
	(0.012)	(0.180)	(0.005)	(0.008)	(0.016)	(0.008)	(0.004)	(0.006)
% High sch, grads[a]	0.012	0.162	−0.003	−0.007	−0.013	−0.008	0.001	0.007
	(0.012)	(0.126)	(0.005)	(0.008)	(0.015)	(0.008)	(0.004)	(0.006)
% Minorities	0.081***	0.420**	0.032***	0.053***	0.065***	0.050***	0.030***	0.035***
	(0.008)	(0.173)	(0.004)	(0.006)	(0.010)	(0.006)	(0.003)	(0.004)
% Hispanic	0.119***	1.143***	0.060***	0.091***	0.104***	0.082***	0.059***	0.061***
	(0.016)	(0.237)	(0.008)	(0.012)	(0.021)	(0.011)	(0.006)	(0.008)
% Female	0.254***	0.968	0.096***	0.152***	0.213***	0.140***	0.101***	0.112***
	(0.054)	(0.761)	(0.017)	(0.027)	(0.051)	(0.026)	(0.015)	(0.021)
% Age 20 to 29[b]	0.329***	1.817	0.211***	0.313***	0.431***	0.296***	0.189***	0.235***
	(0.092)	(1.822)	(0.035)	(0.056)	(0.098)	(0.053)	(0.026)	(0.042)
% Age 30 to 39[b]	0.080	−3.622**	0.042*	0.083**	0.016	0.071*	0.040*	0.044
	(0.068)	(1.543)	(0.025)	(0.042)	(0.071)	(0.039)	(0.022)	(0.028)
% Age 40 to 49[b]	0.682***	4.736**	0.340***	0.527***	0.553***	0.462***	0.294***	0.375***
	(0.129)	(2.014)	(0.047)	(0.074)	(0.134)	(0.071)	(0.034)	(0.057)
% Age 50 to 59[b]	0.067	1.891	0.126***	0.195***	0.221**	0.180***	0.135***	0.126***
	(0.086)	(1.701)	(0.038)	(0.062)	(0.112)	(0.059)	(0.032)	(0.041)
% Age 60 and over[b]	0.220***	1.126	0.143***	0.247***	0.198***	0.224***	0.136***	0.162***
	(0.063)	(1.137)	(0.023)	(0.036)	(0.064)	(0.034)	(0.018)	(0.027)
Household income	0.018	−0.324	0.024***	0.029***	0.072***	0.031***	0.017***	0.026***
	(0.011)	(0.290)	(0.004)	(0.007)	(0.013)	(0.007)	(0.003)	(0.005)

Table A3: (continued)

	Per cap Not logged (1)	Not per cap Not logged (2)	Not per cap Logged (3)	Log per cap Between 7–1 and 9–1 (4)	Log per cap Current (5)	Log per cap Between 7 and 1 and current (6)	Log per cap Mask always only (7)	Log per cap Mask always, freq., and sometimes (8)
Population density	-1.61×10^{-4}***	4.67×10^{-4}	-4.13×10^{-5}***	-8.61×10^{-5}***	-7.60×10^{-5}**	-7.88×10^{-5}***	-3.06×10^{-5}***	-4.41×10^{-5}***
	(5.56×10^{-5})	(1.68×10^{-3})	(1.25×10^{-5})	(2.01×10^{-5})	(3.39×10^{-5})	(1.82×10^{-5})	(1.09×10^{-5})	(1.33×10^{-5})
1st stage F-stat of excl. IV	68.22	73.33	49.27	49.31	49.73	49.73	84.93	43.29

All estimations also include a constant. Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. [a]The reference group is % with no high school degree. [b]The reference group is % under the age of 20.

Table A4: Reproducing main results with COVID-19 cases. (Dependent variable: Log per capita COVID cases on and before September 1.)

	OLS	**OLS**	**OLS**	**OLS**	**IV**	**IV**
	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**
Mask usage frequently/always	0.004***	−0.008***	−0.009***	−0.013***	−0.020***	−0.103***
	(0.002)	(0.001)	(0.001)	(0.001)	(0.003)	(0.015)
		0.613***	0.604***	0.492***	0.620***	0.498***
Log per cap. COVID deaths July 1		(0.010)	(0.010)	(0.011)	(0.010)	(0.021)
		−0.008***	−0.009***	−0.013***	−0.020***	−0.103***
Population		✓	✓	✓	✓	✓
Log all deaths per capita 2016			✓	✓	✓	✓
Other controls				✓		✓
F-test of exclud. instr. in 1st stage					434.58	47.59
Hausman-WU-Durbin F-stat p-value					0.0000	0.0000

All estimations also include a constant. The estimated coefficients for other controls can be found in Table A2. Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively.

Table A5: Including social distancing measure as a control.

	Mask usage as left-hand side variable (1st stage)	COVID-19 deaths as left-hand side variable (2nd stage)
Mask usage		−0.086***
		(0.012)
% Vote Trump	−0.170***	
	(0.020)	
Social distancing	0.109	0.038*
	(0.219)	(0.022)
Log per cap. COVID deaths July 1	0.700***	0.751***
	(0.182)	(0.020)
Population per 100,000	−0.276***	−0.029***
	(0.051)	(0.006)
Log per cap. All deaths 2016	2.300***	0.296***
	(0.214)	(0.037)
% College[a]	−0.010	−0.003
	(0.046)	(0.005)
% High school graduates[a]	−0.020	−0.002
	(0.046)	(0.005)
Percentage minorities	0.097***	0.028***
	(0.018)	(0.003)
Percentage hispanic	0.370***	0.050***
	(0.020)	(0.006)
Percentage female	0.575***	0.084***
	(0.125)	(0.015)
Percentage age 20 to 29[b]	1.799***	0.185***
	(0.130)	(0.028)
Percentage age 30 to 39[b]	0.271	0.032
	(0.225)	(0.023)
Percentage age 40 to 49[b]	2.617***	0.304***
	(0.218)	(0.035)
Percentage age 50 to 59[b]	1.524***	0.106***
	(0.219)	(0.032)
Percentage age 60 and over[b]	1.184***	0.127***
	(0.088)	(0.017)
Average household income (1000s)	0.221***	0.019***
	(0.025)	(0.004)
Population density (square miles)	-3.54×10^{-4}***	-3.41×10^{-5}***
	(8.05×10^{-5})	(1.06×10^{-5})
F-stat of the excluded instrument	72.08	
Number of observations	2766	2766

All estimations also include a constant. Heteroskedasticity-robust standard errors are in parentheses. *, **, and *** represent significance at the 10%, 5%, and 1% levels, respectively. [a]The reference group is % with no high school degree. [b]The reference group is % under the age of 20.

References

Bahl, P., S. Bhattacharjee, C. de Silva, A. A. Chughtai, C. Doolan, and C. R. MacIntyre. 2020. "Face Coverings and Mask to Minimise Droplet Dispersion and Aerosolisation: A Video Case Study." *Thorax* 75 (11): 1024–5.

Bundgaard, H., J. S. Bundgaard, D. E. T. Raaschou-Pedersen, C. von Buchwald, T. Todsen, J. B. Norsk, and K. Fogh. 2020. "Effectiveness of Adding a Mask Recommendation to Other Public Health Measures to Prevent SARS-CoV-2 Infection in Danish Mask Wearers: a Randomized Controlled Trial." *Annals of Internal Medicine* 174: 335–43.

Chernozhukov, V., H. Kasaha, and P. Schrimpf. 2020. "Causal Impact of Masks, Policies, Behavior on Early COVID-19 Pandemic in the US." arXiv preprint arXiv:2005.14168.

IHME COVID-19 Forecasting Team. 2020. "Modeling COVID-19 Scenarios for the United States." *Nature Medicine* 27: 94–105.

Karaivanov, A., S. E. Lu, H. Shigeoka, C. Chen, and S. Pamplona. 2020. *Face Masks, Public Policies and Slowing the Spread of COVID-19: Evidence from Canada (No. w27891)*. Cambridge, Massachusetts: National Bureau of Economic Research.

Lee, D. L., J. McCrary, M. J. Moreira, and J. Porter. 2020. "Valid T-Ratio Inference for IV." arXiv preprint arXiv:2010.05058.

Lindsley, W. G., F. M. Blachere, B. F. Law, D. H. Beezhold, and J. D. Noti. 2020. "Efficacy of Face Masks, Neck Gaiters and Face Shields for Reducing the Expulsion of Simulated Cough-Generated Aerosols." medRxiv.

Lyu, W., and G. L. Wehby. 2020. "Community Use of Face Masks and COVID-19: Evidence from A Natural Experiment of State Mandates in the US: Study Examines Impact on COVID-19 Growth Rates Associated with State Government Mandates Requiring Face Mask Use in Public." *Health Affairs* 39 (8): 1419–25.

Machado, J. A., and J. S. Silva. 2019. "Quantiles via Moments." *Journal of Econometrics* 213 (1): 145–73.

Mitze, T., R. Kosfeld, J. Rode, and K. Wälde. 2020. "Face Masks Considerably Reduce COVID-19 Cases in Germany: A Synthetic Control Method Approach." *Proceedings of the National Academy of Sciences of the U S A* 117: 32293–301.

Staiger, D., and J. H. Stock. 1997. "Instrumental Variables Regression with Weak Instruments." *Econometrica* 65 (3): 557–86.

Stock, J., and M. Yogo. 2005. "Asymptotic Distributions of Instrumental Variables Statistics with Many Instruments." In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*, pp. 109–20.

Verman, S., M. Dhanak, and J. Frankenfield. 2020. "Visualizing the Effectiveness of Face Masks in Obstructing Respiratory Jets. June 2020." *Physics of Fluids* 32 (6): 061708.

Yilmazkuday, H. 2020. "Fighting against COVID-19 Requires Wearing a Face Mask by Not Some but All." Available at SSRN 3686283.

Zhang, J., J. Li, T. Wang, S. Tian, J. Lou, X. Kang, and Y. Chen. 2020. "Transmission of SARS-CoV-2 on Aircraft." Available at SSRN 3586695.

Data Sources

Death data and county population: https://usafacts.org/visualizations/coronavirus-COVID-19-spread-map/.

CDC Overall Mortality Data: https://wonder.cdc.gov/wonder/help/cmf.html#, https://wonder.cdc.gov/cmf-icd10.html.

The New York Times and Dynata Mask-Wearing Survey Data: https://github.com/nytimes/COVID-19-data/tree/master/mask-use.

County level presidential election data: https://dataverse.harvard.edu/dataset.xhtml?persistentId=10.7910/DVN/VOQCHQ.

Social Distancing Data: https://www.unacast.com/post/the-unacast-social-distancing-scoreboard.