The protective effect of smoking against COVID-19: A population-based study using instrumental variables

Zakharov, Nikita

University of Wisconsin–Madison

19 June 2020

The protective effect of smoking against COVID-19: A population-based study using instrumental variables

Nikita Zakharov*

University Wisconsin-Madison

June 20, 2020

Abstract

A low prevalence of smokers among the confirmed patients with COVID-19 has been reported by multiple hospital-based studies, and this observation gave rise to a hypothesis that smoking has a protective effect against the novel coronavirus. We test this prediction in a population-based study across the US states and use an instrumental variable approach to address the endogeneity of smoking rates. We find that a higher prevalence of smoking has a significant negative effect on the spread and the severity of the COVID-19 pandemic across the US state: it decreases the per capita number of registered cases, the case fatality rate, and the excess mortality. The protective effect is more pronounced in subgroups of the population that are more likely to be smokers: men of all ages and females of the older cohort. Our findings are robust to the inclusion of a broad range of control variables, exclusion of outliers, and placebo tests. Despite the protective effect against the COVID-19, smoking remains detrimental for health in the long-term, and we show that states with a higher rate of smoking also have higher mortality in the year before the outbreak.

Keywords: COVID-19, US states, smoking, cigarettes, case fatality rate, excess mortality

JEL codes: H71, I12, I18,

---

* Nikita Zakharov, University Wisconsin-Madison, Taylor Hall 413, Lorch Street, Madison 53706, email: zakharov@wisc.edu. I declare no conflict of interest.

1 Introduction

This paper investigates a major puzzle posed by the global pandemic regarding a potentially protective effect of smoking against the novel coronavirus decease (COVID -19). From the very start of the outbreak, smoking was considered to be a major risk factor both for infection spread and severity of the COVID-19 illness (e.g., Brake et al. 2020). This expectation was rooted in the previous research that finds smokers to be significantly more vulnerable to the respiratory deceases in general (e.g., Bello et al. 2014). However, it was soon observed that the prevalence of smokers hospitalized with COVID-19 was considerably lower compared to the smoking rate in the general population. This finding was confirmed using data on hospitalized patients in China (Arsalan's et al. 2020a), the United States (Chow et al. 2020; Petrilli et al. 2020), France (Miyara et al. 2020), and Italy (Gonzalez-Rubio et al. 2020). The difference in the prevalence of smokers hospitalized with COVID-19 is not only consistent across countries but also substantial in its magnitude: for example, Farsalinos et al. (2020a) estimate it to be approximately one-fourth of the expected gender-adjusted smoking prevalence in China.[1] If correct, this regularity would imply a protective effect of smoking against COVID-19 that might be the key to developing an effective therapy for the decease. Yet, the current evidence is grounded exclusively in hospital-based studies that suffer from limitations such as data quality, sampling bias, and, consequently, low external validity, and further research requires well-designed population-based studies (WHO 2020).

We answer this call by providing an econometric study of the relationship between the spread and mortality from COVID-19 and smoking across the US states. The USA represents about 30% of the world's registered COVID-19 cases and deaths. The ample heterogeneity in the spread and severity of the pandemic across the states provides a unique empirical setting for our analysis. We evaluate the relationship between the three main statistics of the COVID-19 epidemic, such as the number of infections, the case fatality rate (CFR), and the excess mortality, and smoking prevalence across the US states. Because smoking rates in the

[1] The more recent study by Farsalinos et al. (2020b) includes additional data from the USA and South Korea and finds the prevalence of smokers to be even smaller: about one-fifth of prevalence in corresponding national population.

population are endogenous, we apply the instrumental variables approach (IV). We take advantage of the state-to-state diffusion of anti-smoking policies (e.g., French and Popovici 2011) and show that higher past state-taxes on cigarettes in the neighboring states are a strong predictor of lower rates of smoking and cigarette consumption. Our main findings are the significant negative effects of smoking on the spread of the COVID-19 and its fatality in terms of the CFR and population fatality rate (PFR) approximated by the excess mortality. The magnitude of the effect is substantial: 1 percentage point increase in smoking leads to a reduction in infection incidents, the CFR, and the PFR by 0.3, 1, and 0.8 percentage points, respectively. We further look at the age and gender composition of deaths from COVID-19 and find that the protective effect is robust among the adult male population irrespective of age. In the female population, the protective effect of a similar size is detected only for women older than 65 years, and there is no effect for women younger than 65 because of a relatively lower smoking rate in this age group (Department of Health and Human Services 2001). Despite the protective effect against the novel coronavirus, smoking is deadly, as we demonstrate in the analysis of the mortality rate across the US state in the year before the pandemic. Finally, our results are robust to the exclusion of virus-hotspot states of New York and New Jersey, controlling for the virus spread in the neighboring states, and using deaths from influenza as a placebo test.

This paper finds evidence that supports the potentially protective effect of smoking against the novel coronavirus, however, the biological mechanism behind this effect is yet to be explored in future research. Currently, the most credible explanation is the anti-inflammatory and positive immune-modulating effect of smoking attributed to nicotine (e.g., Kloc et al. 2020; Polosa and Caci 2020). Nicotine is assumed to make individuals more tolerant and less reactive to the infection, and, as a result, they are less likely to experience a cytokine release syndrome associated with COVID-19-related high mortality (Garufi et al. 2020). An alternative, more surprising explanation is suggested by Changeux et al. (2020), who hypothesize that nicotine competes with the virus for binding with the receptors for entry, thus decreasing the risk of contracting the decease, but this has not been tested in medical research previously.

Our paper contributes to the growing literature on smoking and mortality from COVID-19 by providing the first population-based study with a robust identification strategy that employs a novel instrument for smoking. To our knowledge, Mollalo et al. (2020) is the only paper that includes smoking prevalence in the US states as a control variable for the spread of the novel coronavirus. This paper does control for the endogeneity of smoking and finds no significant results.

Our findings have immediate policy implications. The potential protective effect of smoking must be examined more carefully for the purpose of therapy development. Further research should evaluate the protective effect of smoking and nicotine using randomized-control trials, and the mechanisms behind this effect should be identified and isolated from the direct harm of smoking. Additionally, our findings should be taken into consideration before advocating the cessation of smoking as an effective measure against the novel virus (e.g., Hefler and Gartner 2020). Our results should not be used to promote smoking as it is a direct cause of millions of premature deaths worldwide.

The paper is organized as follows: Section 2 describes the data and the estimation strategy; Section 3 presents the main regression results; Section 4 looks at the deaths per capita among gender and age groups and provides the results for the placebo tests in the respective subgroups using influenza deaths per capita; Section 5 concludes.

2 Data and estimation strategy

The USA is currently a country with the most registered cases and fatalities from the novel coronavirus: over 2 million cases and 113 thousand deaths, which constitutes almost 30% of all recorded cases and deaths around the world. After the declaration of a national emergency by the White House three months ago and the current steady decline of the first wave of the pandemic, it might be the right time to assess the determinants of the COVID-19 severity in the USA.

One of the main features of the pandemic in the USA is an uneven geographical distribution of the cases and mortality from the virus. We evaluate this regional heterogeneity using the three main indicators of the severity of COVID-19 in each state and map them in Figure 1. We use the standard data sources in the COVID-19 literature: Worldometers.info and CDC database. First, Panel A of Figure 1 illustrates the differences in the spread of registered cases of COVID-19 per thousand population as of June 8, 2020. This measure represents mainly the prevalence of symptomatic cases and not the overall rate of infection since the asymptomatic cases are substantially underreported (e.g., Manski and Molinari 2020; Nishiura et al. 2020). The second measure is the CFR in percentage mapped in Panel B. The CFR is a rough approximation of the lethality that may be overstated if actual cases are underreported, however, this limitation is of a lesser concern due to the extensive testing in the USA.[2] An alternative measure of the severity of the COVID-19 is the excess mortality in the population of the state, which is a close approximation to the population fatality rate (PFR). CDC provides the estimates for the excess mortality as the difference between observed mortality between February 1 and June 3 and the expected mortality in the same period based on the historical data.[3] Panel C shows the excess mortality per thousand population. Interestingly, the rate of smoking in the US also varies substantially across the states, as we show in Panel D using the recent survey data from the CDC Behavioral Risk Factor Surveillance System (BRFSS) measured as a percent of the population that smoked in 2018.

Can the uneven character of the pandemic be explained by the differences in smoking prevalence across the states? For instance, among the three states on the West Coast, the state with the highest smoking rates, Oregon, seems to handle the pandemic significantly better than its neighbors with fewer smokers, Washington State and California. Does this pattern hold for all the states? Figure 2 shows the correlation between smoking rate and the regional COVID-19 metrics from Figure 1 using scatterplots with linear trends for 49 of the US states, omitting isolated states of Hawaii and Alaska. Additionally, Figure 2 includes a scatterplot for overall mortality in 2019 to re-confirm smoking as a risk factor for premature deaths in the pre-pandemic year (e.g., Kontis et al. 2015).

---

[2] While the rate of testing was low during the onset of the pandemic, it has been significantly improved shortly after. To this moment, the per capita number of tests in the USA has become one of the highest worldwide.

[3] The methodology for calculating the excess mortality and the data are available at the CDC website: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm.

Figure 1: Differences in COVID-19 severity and smoking rates across the US states

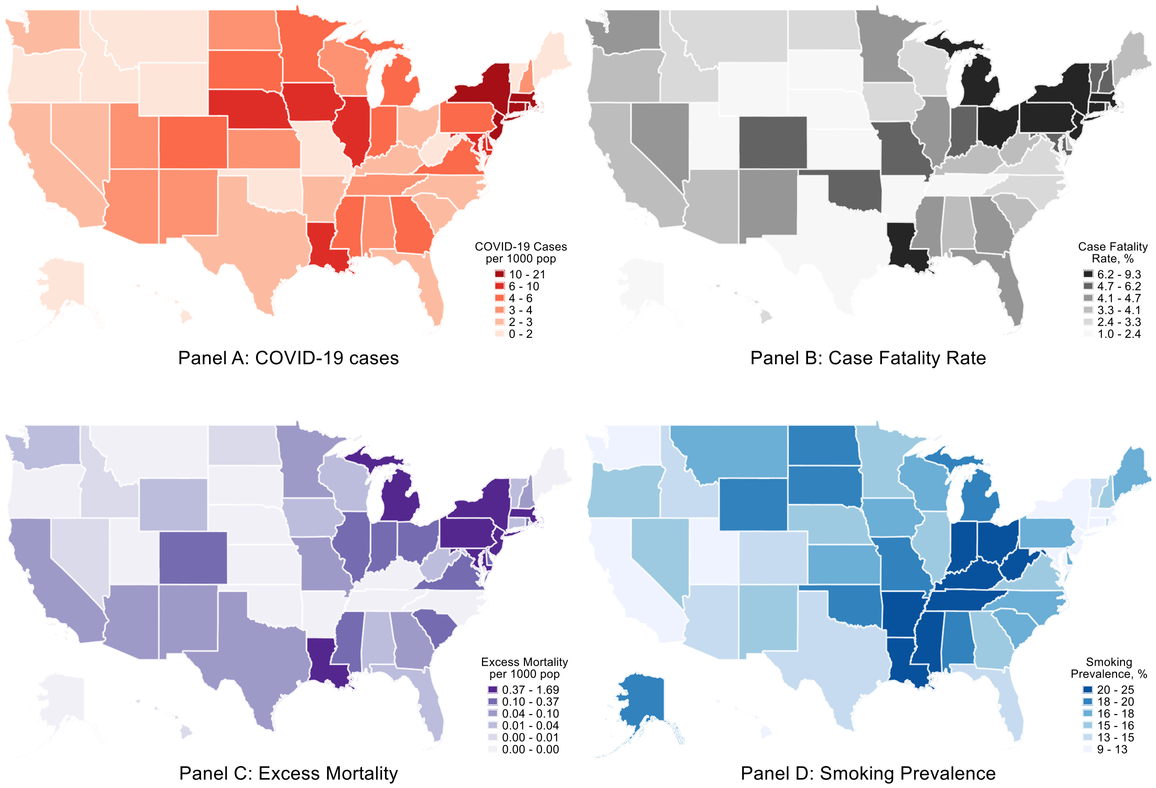
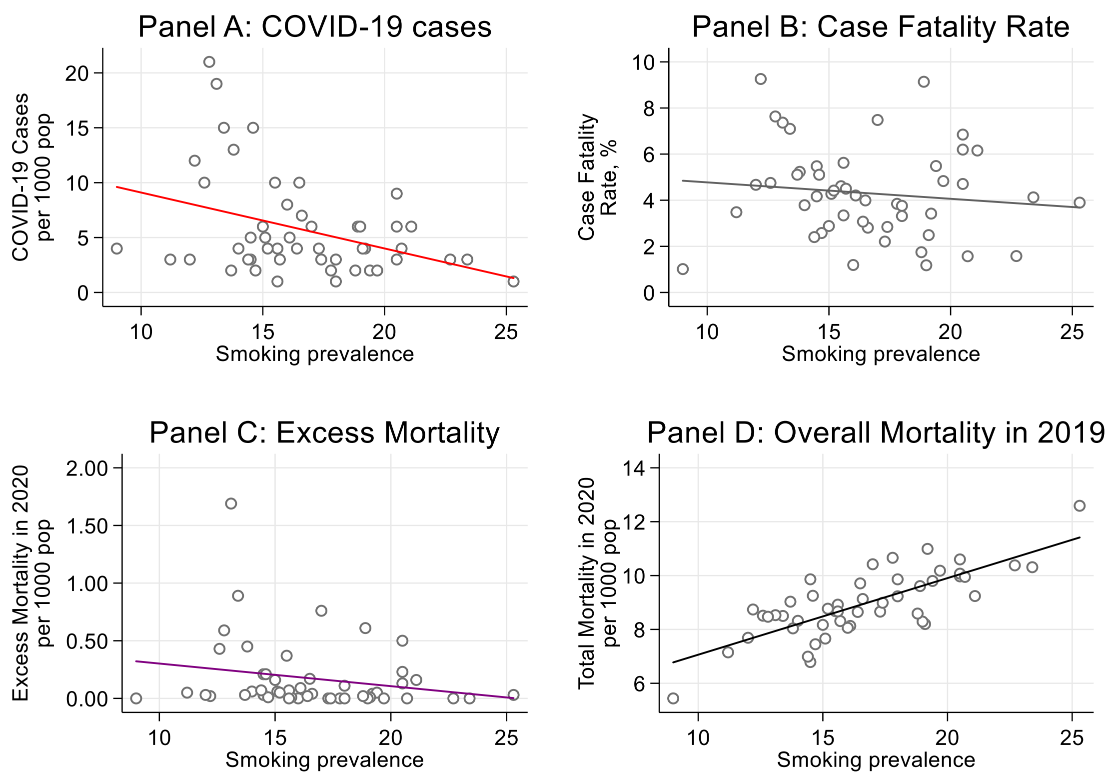


Figure 2: Correlations between the COVID-19 severity and smoking prevalence



Panel A of Figure 2 shows a negative relationship between smoking rates and the number of cases of COVID-19 in line with the empirical evidence from the hospital-based studies. This could be interpreted as fewer overall cases of infection due to the protective effect or, most likely, fewer symptomatic cases if people who got sick but did not experience any complications refrained from being tested. Panel B shows a negative relationship between smoking and the CFR: there are fewer deaths among the registered cases. This potentially means that smoking decrease not only the number of symptomatic cases but also the likelihood of those cases to have a fatal outcome. Penal C looks at a different measure of fatalities from COVID-19 among the population and demonstrates a negative trend: states with fewer smokers experienced higher overall human losses. While smoking could be protective against the novel virus, it remains the major risk factor for premature deaths as we show in Panel D: the level of overall mortality in 2019 was higher in states with more prevalent smoking.

While illustrative, correlation does not imply causation. We proceed further with the regression analysis that addresses the endogeneity of smoking and includes the controls for the social and economic characteristics of the states.

Our estimation strategy is the following:

$$COVID19_i = \alpha + \beta\ Smoking_i + X_i + \varepsilon_i, \qquad (1)$$

where i indicates the state; $COVID19_i$ is the spread or the severity of the decease measured by the number of registered cases of COVID-19 per million population in logarithms (*Covid_Cases*), the CFR in percentage (*Covid_CFR*), and a logarithm of the number of excess deaths per million population plus one (*Excess_Deaths*); *Smoking* indicates the prevalence of smoking in percentage; X_i is the vector of control variables; and ε_i is an error term.

The data for the first two outcome variable, *Covid_Cases,* and *Covid_CFR*, come from Worldometers.info, which collects and double-checks the data from federal and subnational sources such as CDC and state health departments.[4] It was collected on June 8, 2020. The data for excess mortality are provided by the CDC. The most recent estimates are available for the period from February 1 until June 3.[5]

[4] The methodology is available at the website of Worldometers.info: https://www.worldometers.info/coronavirus/about/.

[5] The methodology is available at the CDC website: https://www.cdc.gov/nchs/nvss/vsrr/covid19/excess_deaths.htm.

We construct three variables for smoking prevalence. The first two variables are the share of the population who smokes (*Smoking*) and the share of the population who smokes daily (*Smoking_*Daily). The data come from the CDC BRFSS representative survey that has interviewed over half a million Americans in 2018. Another variable for smoking is the annual cigarette consumption per capita, measured as a logarithm of the number of packs purchased in the state per capita (*Cigarettes*). The data are from the CDC (2018). While cigarette consumption is a less precise measure of smoking rates, we use it to show the robustness of our results.

We address the endogeneity of smoking by an instrumental variable approach. Our instrument for smoking is motivated by an adverse effect of the state-taxes on cigarette prices, and consequently, cigarette consumption. While state-taxes are a common instrument for smoking (French and Popovici 2011), they can be themselves endogenous to the state characteristics or other state policies. To construct the exogenous instrument, we exploit the existence of the state-to-state diffusion of anti-smoking policies (Shipan and Volden 2006). We hypothesize that the current cigarette state-tax as a type of anti-smoking policy is affected by the past levels of cigarette taxes in the neighboring states. The cigarette taxes of the neighboring states from the previous years are orthogonal to the current characteristics of the state and influence the smoking rate only via the regional convergence process of cigarette taxes and, therefore, our instrument satisfies the exclusion restrictions.

Our instrument (*Neighbor_ Tax*) is constructed as an average of state-taxes for a pack of cigarettes in the neighboring states calculated for the year 2013 using CDC (2018) data.[6] This allows five years for the tax convergence across the states before our most recent regional estimates of smoking. The convergence over this period is considerable: the difference between the cigarette tax in the state and its neighbors was 30% smaller in 2018 compared to the same difference in 2013.

We show the effect of our instrument on smoking and cigarette consumption in the following regression model:

$$Smoking_i = a + \lambda \, Neighbor\_Tax_{i_{2013}} + e_{i,} \qquad (2)$$

[6] Our results are not sensitive to the choice of the year as we show further in a robustness check.

where i=1,...,49 indexes the US states, *Smoking* is the smoking rate in state i, *Neighbor_Tax* is the average state tax on a pack of cigarettes in the year 2013, and e is an error term.

The control variables include a range of demographic, economic, and political characteristics of the US states. We control for the logarithm of the total population (*Population*), population density (*Density*), the share of the population older than 65 years as this group has the highest risk of severe outcomes from COVID-19 (*Population_65*), the share of the population older than 45 years and younger than 65 since this group also has a notable fatality rate for COVID-19 (*Population_45*).[7] The additional demographic factor is the share of the white population (*White*) since a few early studies warn about the observed higher risk of death among ethnic minorities (e.g., Millett et al. 2020). Among the economic factors, we include the average annual per capita income adjusted for the cost of living (*Income*), inequality measured by the Gini coefficient (*Inequality*), poverty as a share of population under the federal poverty line (*Poverty*). The data from 2018 CDC BRFSS survey allows adding health-related factors that might affect the severity of COVID-19: overall healthiness measured as a share of people who reported their health condition as excellent (*Healthy*), and the proportion of respondents with body mass index that exceeds 25 (*Overweight*) since being overweight or obese substantially increases the risk of developing complications from COVID-19 (e.g., Cai et al. 2020). The non-pharmaceutical interventions were aimed at reducing the spread of coronavirus, and, therefore, we control for the overall effect of lockdowns by using the average daily percentage of the population that were staying at home (*Stayhome*). The *Stayhome* variable is constructed from the mobile device location database by Zhang et al. (2020). It is an actual estimate of the population that has reduced its mobility between March 1 and May 31. Finally, we add the party affiliation of the governor (*Gov_Republican*) if the Republican leadership could have affected the local response to the virus outbreak.[8]

The summary of variables and the data sources is presented in Appendix A, Table A.1.

[7] The population younger than 45 years consitute about 2.5% of all the fatalities, while the population between 45 and 64 years old represent 17% of fatalities, and people older than 65 consititute 80.5% of all fatalities respectively.

[8] For a discussion one can look into the article of Wall Street Journal available at: https://www.wsj.com/articles/why-coronavirus-increasingly-exacerbates-the-red-blue-divide-11589810871

3 Results

Our estimation strategy is based on the variation in smoking prevalence due to our instrument. Table 1 presents the first stage of the IV estimation following equation (2) for each proxy of smoking, without any control variables and with controls specified in our main estimation equation (1). It shows that the average of past cigarette taxes in the neighboring states in 2013, *Neighbor_Tax,* is a strong predictor of a smaller share of active smokers (column 1-2), including persistent smokers (column 3-4), and lower cigarette consumption (column 5-6) in 2018. If the neighboring states had the cigarette tax higher by one dollar in 2013, that would result in 1.41 percentage points reduction in smoking rate in 2018. The effect is highly significant in estimations with (columns 2, 4, 6) and without control variables (columns 1, 3, 5). The F-statistics of the first stage above ten indicates strong instruments (Staiger and Stock 1997).

Table 1: Past state-taxes on cigarettes in the neighboring states and the current prevalence of smoking

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	Smoking		Smoking_Daily		Cigarettes	
Neighbor_Tax	-2.017***	-1.405***	-1.588***	-1.179***	-0.210***	-0.229***
	(-5.42)	(-4.05)	(-4.93)	(-4.25)	(-2.87)	(-3.25)
Controls	No	Yes	No	Yes	No	Yes
R squared	0.19	0.88	0.15	0.89	0.11	0.76
F-statistics	29.42	16.38	24.3	18.04	8.26	10.58
LM test	0.002	0.001	0.003	0.001	0.026	0.004

Notes: t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Number of states N=49

Table 2 presents the main results of our IV estimation for the three indicators of the severity of COVID-19 and the overall mortality rate in four panels in an order similar to Figure 2. Panel A shows that smoking rates have a negative and statistically significant effect on the spread of COVID-19 cases in the states. This effect is consistent across all three variables for smoking with and without controls. The magnitude suggests that an increase in *Smoking* by one standard deviation would reduce the number of registered infections by 1 percentage points. The changes by one standard deviation in *Smoking_Daily* or *Cigarettes* yield almost the same magnitude of the effect on COVID-19 outcomes.

Table 2: The effect of smoking prevalence on the severity of COVID-19 across the US states, IV

Panel A: Registered cases of COVID-19 per million population, in logs						
	(1)	(2)	(3)	(4)	(5)	(6)
Smoking	-0.229***	-0.301***				
	(-2.97)	(-3.27)				
Smoking_Daily			-0.291***	-0.359***		
			(-3.06)	(-3.33)		
Cigarettes					-2.203***	-1.851***
					(-2.74)	(-2.86)
Controls	No	Yes	No	Yes	No	Yes
Panel B: COVID-19 Case fatality rate, percent						
	(1)	(2)	(3)	(4)	(5)	(6)
Smoking	-0.782***	-1.001***				
	(-4.53)	(-2.83)				
Smoking_Daily			-0.993***	-1.194***		
			(-4.37)	(-2.92)		
Cigarettes					-7.510***	-6.153***
					(-2.95)	(-2.79)
Controls	No	Yes	No	Yes	No	Yes
Panel C: Excess deaths per million population plus one, in logs						
	(1)	(2)	(3)	(4)	(5)	(6)
Smoking	-0.614***	-0.842**				
	(-2.84)	(-2.13)				
Smoking_Daily			-0.779***	-1.004**		
			(-2.85)	(-2.30)		
Cigarettes					-5.891**	-5.175**
					(-2.34)	(-1.98)
Controls	No	Yes	No	Yes	No	Yes
Panel C: Overall mortality in 2019 per million population, in logs						
	(1)	(2)	(3)	(4)	(5)	(6)
Smoking	0.01	0.027***				
	(1.36)	(4.26)				
Smoking_Daily			0.013	0.032***		
			(1.38)	(4.06)		
Cigarettes					0.099	0.163***
					(1.36)	(2.89)
Controls	No	Yes	No	Yes	No	Yes

Notes: t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Number of states N=49.

Panel B in Table 2 shows the effect of smoking on the CFR. The higher smoking rate is negatively associated with fatality among the registered cases. One standard deviation increase in *Smoking, Smoking_Daily,* and *Cigarettes* would decrease the CFR by 3.3, 3, and 2.9 percentage points, respectively. It is important to acknowledge that our estimates are local average treatment effects and, thus, they can be larger than the average treatment effect due to the heterogeneity in the treatment. The magnitude is striking since smoking reduces not only the number of symptomatic cases as has been established in Panel A, but also the share of fatalities among these cases.

The excess mortality is an important indicator of the overall toll of the pandemic across the states. In Panel C, we find that smoking is associated negatively with the rate of excess mortality. The magnitude of the one standard deviation increase in *Smoking, Smoking_Daily,* and *Cigarettes* leads to a reduction in the excess mortality by 2.8, 3, and 2.4 percentage points, respectively, which is very similar to the effect size of smoking prevalence on the CFR.

Finally, Panel D shows that our variables for smoking prevalence have a positive effect on overall mortality in the year before the pandemic, but the effect is statistically significant only after we include controls. One standard deviation increase in *Smoking* is associated with 0.1 percentage points increase in overall mortality. The finding is in line with the existing consensus that smoking causes premature death (e.g., Nam and Rodgers 1998).

There are a few interesting correlations between the control variables and the severity of COVID-19 that can be observed in Appendix A, Table A.2. For example, COVID-19 is relatively more widespread in states with a higher density of population, higher average income, and higher poverty rates. The positive correlation between the share of the white population and the spread of the infection and its lethality is in line with the findings of Egorov et al. (2020) that diverse societies are more likely to follow social distancing guidelines in the times of COVID-19 pandemic. A more healthy and less overweight population has a lesser spread of the virus, but these parameters are not associated with the fatality rates. Surprisingly, the percent of inhabitants staying at home during the pandemic is strongly associated with more cases and deaths from COVID-19, which might be explained by reverse causality when the population conforms more to the social distancing in relatively more troubled locations.

Our estimation results are robust to several alternative specifications. First, Table A.3 in Appendix A shows that results do not change if we omit the states of New York and New Jersey as potential outliers since they represent the largest outbreaks in the country. Next, we test if controlling for the spread and the death toll in the neighboring states affects the results. For this purpose, we construct two control variables: one is the logarithm of the average number of cases per million population in the neighboring states (*Neighbor_Cases*), and another control is the excess mortality per million population in logarithms (*Neighbor_Excess_Deaths).* Table A.4 in Appendix A shows that the number of infections in the neighboring states is positively associated with the number of registered cases in the state, but this does not change our main results for smoking. Finally, we test if our results are sensitive to the choice of the year for constructing our instrument. We run the same IV estimation using the average cigarette state-taxes from 2010 and present the results in Table A.5 in Appendix A. The coefficients have the same sign and similar magnitudes as in our main estimations. We confirm the robustness of the protective effect of smoking against COVID-19 once again.

4 Smoking and the fatality of COVID-19 across gender and age

This section investigates the effect of smoking across the gender and age groups. We take advantage of the provisional data of confirmed deaths from COVID-19 assembled by the CDC, which covers the period from February 1 until June 3, 2020. We expect the protective effect of smoking against the virus to be more robust among the male population and less consistent among females, especially in the younger age group, since they are less likely to smoke.

We aggregate data on all the registered deaths from COVID-19 in four groups by gender and by two age categories: older than 65 and older than 45 and younger than 65. We focus only on the population older than 45 years old since most (97.5%) of the deaths in the USA occur among this category. We divide each group by the population in the state of the respective gender and age and take the logarithm.[9] The outcome measures can be interpreted as a fatality rate in the respective subgroup of the population.

Table 3 provides the results for the mortality among gender and age categories. We find the statistically significant protective effect of smoking among both age groups of males (Panel A) and older females (Panel B, columns 4-6), but no significant effect among females aged between 45-64. The heterogeneity can be explained by a higher proportion of smokers among those subgroups. While the gender difference in smoking rates among the older cohorts are relatively small (e.g., Waldron 1991; John et al. 2005), younger cohorts of females smoke significantly less than older women or men in both groups (e.g., Burns et al. 1997).

Table 3: Smoking and COVID-19 mortality by age group and gender

Panel A: Male population

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	Male COVID-19 deaths 45-64			Male COVID-19 deaths 65+		
Smoking	-0.374**			-0.425***		
	(-2.30)			(-2.91)		
Smoking_Daily		-0.446**			-0.507***	
		(-2.41)			(-3.11)	
Cigarettes			-2.300**			-2.613***
			(-2.47)			(-2.67)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Female population

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	Female COVID-19 deaths 45-64			Female COVID-19 deaths 65+		
Smoking	-0.287			-0.430**		
	(-1.34)			(-2.48)		
Smoking_Daily		-0.342			-0.513**	
		(-1.38)			(-2.56)	
Cigarettes			-1.763			-2.641**
			(-1.38)			(-2.40)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Number of states N=49.

[9] The data on the registered COVID-19 cases cannot be disaggregated by gender and age, and thus we are unable to construct the CFR in the respective subgroup of the population.

The CDC's data allows us to perform a placebo test using the fatalities from influenza for the same time period. The main assumption is that smoking does not exhibit any protective effect in cases of other respiratory deceases such as influenza and can be even potentially harmful (e.g., Bello et al. 2014; Lawrence et al. 2019). We assemble the fatality rate for influenza across both genders and the same age groups as in Table 4.

The main results in Table 4 suggest that smoking prevalence is not associated with lower levels of fatalities from influenza. On the contrary, all the coefficients are positive, which is more in accord with the potentially harmful effect of smoking for respiratory deceases, however, they are not statistically significant.

Table 4: Smoking and influenza mortality by age group and gender

Panel A: Male population

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	Male influenza deaths 45-64			Male influenza deaths 65+		
	(0.74)			(0.43)		
Smoking_Daily		0.125			0.065	
		(0.75)			(0.43)	
Cigarettes			0.644			0.334
			(0.69)			(0.42)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Panel B: Female population

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Variable:	Male influenza deaths 45-64			Male influenza deaths 65+		
Smoking	0.145			0.047		
	(1.32)			(0.46)		
Smoking_Daily		0.173			0.056	
		(1.27)			(0.46)	
Cigarettes			0.894			0.29
			(1.24)			(0.44)
Controls	Yes	Yes	Yes	Yes	Yes	Yes

Notes: t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Number of states N=49.

5 Conclusion

This paper is a first population-based study that investigates the effect of smoking prevalence on the spread and the severity of novel coronavirus in the population of the US states. Our main finding is the persistent negative relationship between smoking rates and the number of registered cases of COVID-19, the CFR, and excess mortality. Our estimation strategy accounts for the endogeneity of smoking rates by using past levels of cigarette state-taxes in the neighboring states as an instrumental variable, suggesting the causality of the effect. We also control for various demographical, economic, and political characteristics of the states and perform multiple robustness tests.

The protective effect of smoking against COVID-19 is puzzling because today, smoking has been consistently proven to be detrimental to individual health in the long-term, and this paper also shows that smoking is positively associated with additional mortality. However, for most of history, tobacco and smoking have been considered therapeutic and widely used in medical practices. Charlton (2004) provides an elaborate historical account of medical uses of tobacco since its adoption from the Native American population who considered the herb to be the potent remedy against the common cold and fever. In 18th and 19th century Europe, tobacco was extensively used against a broad spectrum of infectious deceases as the properties of the nicotine were believed to be antiseptic. Therefore, it was not surprising that smoking was prescribed by doctors during the massive outbreaks of infections, such as the London plague or smallpox outbreak in Bolton (Charlton 2004: 295). Charlton concludes her article with a suggestion that "*we should set aside the prejudices generated by the ill-effects of tobacco smoking and examine the leaves systematically for substances of therapeutic value.*" In the times of the COVID-19 pandemic, the persistent evidence for the protective effect of smoking makes this suggestion extremely relevant.

The purpose of this paper is to provide systematic evidence that would channel the efforts of the scientific community for battling the pandemic into investigating the potential source of the protective effect of smoking and using it for developing potential treatments against COVID-19 and any future coronaviruses.

References

Bello, S., Menéndez, R., Antoni, T., Reyes, S., Zalacain, R., Capelastegui, A., Aspa, J., Borderías, L., Martin-Villasclaras, J. J., Alfageme, I., De Castro, F. R., Rello, J., Luis, M., & Ruiz-Manzano, J. (2014). Tobacco smoking increases the risk for death from pneumococcal pneumonia. *Chest*, *146*(4), 1029-1037. https://doi.org/10.1378/chest.13-2853.

Brake, S. J., Barnsley, K. I., Lu, W., McAlinden, K. D., Eapen, M. S., & Sohal, S. S. (2020). Smoking upregulates angiotensin-converting enzyme-2 receptor: A potential adhesion site for Novel Coronavirus SARS-CoV-2. *Journal of Clinical Medicine*, *9*(3), 1-7. https://doi.org/10.3390/jcm9030841.

Burns, D. M., Lee, L., Shen, L. Z., Gilpin, E., Tolley, H. D., Vaughn, J., & Shanks, T. G. (1997). Cigarette smoking behavior in the United States. In: Burns, D.M., Garfinkel, L., and Samet, J. (Eds.) *Changes in cigarette-related disease risks and their implication for prevention and control. Smoking and tobacco control monograph No. 8.* Bethesda, MD: Cancer Control and Population Sciences, National Cancer Institute, US National Institutes of Health. 13-42.

Cai, Q., Chen, F., Wang, T., Luo, F., Liu, X., Wu, Q., He, Q., Wang, Z., Liu, Y., Liu, L., Chen, J. & Xu, L. (2020). Obesity and COVID-19 severity in a designated hospital in Shenzhen, China. *Diabetes Care*, forthcoming. https://doi.org/10.2337/dc20-0576.

CDC. (2018). State Tobacco Activities Tracking and Evaluation (STATE) system. Centers for Disease Control and Prevention. Online: http://www.cdc.gov/statesystem.

Changeux, J. P., Amoura, Z., Rey, F., & Miyara, M. (2020). A nicotinic hypothesis for Covid-19 with preventive and therapeutic implications. *Qeios PrePrint:* ID:FXGQSB.2. https://doi.org/10.32388/FXGQSB.2.

Charlton, A. (2004). Medicinal uses of tobacco in history. *Journal of the Royal Society of Medicine*, *97*(6), 292-296. https://doi.org/10.1177/014107680409700614.

Chow N, Fleming-Dutra K, Gierke R, Hall A (2020) Preliminary estimates of the prevalence of selected underlying health conditions among patients with coronavirus disease 2019 - United States. February 12 - March 28, 2020. MMWR Weekly Report 69:382–386. https://doi.org/10.15585/mmwr.mm6913e2.

Department of Health and Human Services. (2001). Women and smoking: A report of the Surgeon General. Rockville, MD: US Department of Health and Human Services, Public Health Service, Office of the Surgeon General, 2001.

Egorov, G., Enikolopov, R., Makarin, A., & Petrova, M. (2020). *Divided we stay home: Social distancing and ethnic diversity*. National Bureau of Economic Research, Working paper No. w27277. https://doi.org/10.3386/w27277.

Farsalinos, K., Barbouni, A., & Niaura, R. (2020a). Systematic review of the prevalence of current smoking among hospitalized COVID-19 patients in China: Could nicotine be a therapeutic option? *Internal and Emergency Medicine*, 1-8. https://doi.org/10.1007%2Fs11739-020-02355-7.

Farsalinos, K., Poulas, K., Polosa, R., Barbouni, A., Caponnetto, P., & Niaura, R. (2020c). Prevalence of current smoking and association with adverse outcome in hospitalized COVID-19 patients: A systematic review and meta-analysis. *PrePrints* 2020050113 , May 5. https://doi.org/10.20944/preprints202005.0113.v1.

French, M. T., & Popovici, I. (2011). That instrument is lousy! In search of agreement when using instrumental variables estimation in substance use research. *Health economics*, *20*(2), 127-146. https://doi.org/10.1002/hec.1572.

Garufi, G., Carbognin, L., Orlandi, A., Tortora, G., & Bria, E. (2020). Smoking habit and hospitalization for severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)-related pneumonia: The unsolved paradox behind the evidence. *European Journal of Internal Medicine*, forthcoming. https://doi.org/10.1016/j.ejim.2020.04.042.

Gonzalez-Rubio, J., Navarro-Lopez, C., Lopez-Najera, E., Lopez-Najera, A., Jiménez-Díaz, L., Navarro-López, J. D., & Nájera, A. (2020). What is happening with smokers and COVID-19? A systematic review and a meta-analysis. *Preprints* 2020040540, April 30. https://doi.org/10.20944/preprints202004.0540.v1.

Hefler, M., & Gartner, C. E. (2020). The tobacco industry in the time of COVID-19: Time to shut it down? *Tobacco Control*, *29*, 245-246. http://dx.doi.org/10.1136/tobaccocontrol-2020-055807.

John, U., Hanke, M., Meyer, C., & Schumann, A. (2005). Gender and age differences among current smokers in a general population survey. *BMC Public Health*, *5*(1), 57. http://dx.doi.org /10.1186/1471-2458-5-57.

Kloc, M., Ghobrial, R. M., & Kubiak, J. Z. (2020). How nicotine can inhibit cytokine storm in the lungs and prevent or lessen the severity of COVID-19 infection? *Immunology Letters*, 224, 28-29. https://doi.org/10.1016/j.imlet.2020.06.002.

Kontis, V., Mathers, C. D., Bonita, R., Stevens, G. A., Rehm, J., Shield, K. D., Riley, L. M., Poznyak, V., Jabbour, S., Garg, R. M., Fouad, H., Beaglehole, R., Ezzati, M., & Hennis, A. (2015). Regional contributions of six preventable risk factors to achieving the 25× 25 non-communicable disease mortality reduction target: a modelling study.

Lancet Global Health, *3*(12), e746-e757. https://doi.org/10.1016/S2214-109X(15)00179-5.

Lawrence, H., Hunter, A., Murray, R., Lim, W. S., & McKeever, T. (2019). Cigarette smoking and the occurrence of influenza – Systematic review. *Journal of Infection*, *79*(5), 401-406. https://doi.org/10.1016/j.jinf.2019.08.014.

Manski, C. F., & Molinari, F. (2020). Estimating the covid-19 infection rate: Anatomy of an inference problem. *Journal of Econometrics*, forthcoming. https://doi.org/10.1016/j.jeconom.2020.04.041.

Maryland Transportation Institute (2020). University of Maryland COVID-19 Impact Analysis Platform. University of Maryland, College Park, USA. https://data.covid.umd.edu.

Millett G.A., Jones A. T., Benkeser D., Baral S., Mercer L., Beyrer C., Honermann B., Lankiewicz E., Mena L., Crowley J. S., Sherwood J., & Sullivan P. (2020). Assessing differential impacts of COVID-19 on black communities. *Annals of Epidemiology*, forthcoming. https://doi.org/10.1016/j.annepidem.2020.05.003.

Miyara, M., Tubach, F., Pourcher, V., Morelot-Panzini, C., Pernet, J., & Haroche, J. (2020). Low incidence of daily active tobacco smoking in patients with symptomatic COVID-19. *Qeios PrePrint*, ID: WPP19W.2, April 20. https://doi.org/10.32388/WPP19W.2.

Mollalo, A., Vahedi, B., & Rivera, K. M. (2020). GIS-based spatial modeling of COVID-19 incidence rate in the continental United States. *Science of The Total Environment*, 138884. https://doi.org/10.1016/j.scitotenv.2020.138884.

Nam, C. B., & Rogers, R. G. (1998). Adult mortality differentials associated with cigarette smoking in the USA. *Population Research and Policy Review*, *17*(3), 285-304. https://doi.org/10.1023/A:1005955127544.

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S. M., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A. R., & Linton, N. M. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International Journal of Infectious Diseases.94*, 154–155. https://doi.org/10.1016/j.ijid.2020.03.020.

Petrilli, C. M., Jones, S. A., Yang, J., Rajagopalan, H., O'Donnell, L. F., Chernyak, Y., Tobin, K., Cerfolio, R. J., Francois, F., & Horwitz, L. I. (2020). Factors associated with hospitalization and critical illness among 4,103 patients with COVID-19 disease in New York City. *MedRxiv*, April 11. https://doi.org/10.1101/2020.04.08.20057794.

Polosa, R., & Caci, G. (2020). COVID-19: counter-intuitive data on smoking prevalence and therapeutic implications for nicotine. *Internal and Emergency Medicine*, forthcoming. https://doi.org/10.1007/s11739-020-02361-9.

Shipan, C. R., & Volden, C. (2006). Bottom-up federalism: The diffusion of anti-smoking policies from US cities to states. *American Journal of Political Science*, *50*(4), 825-843. https://doi.org/10.1111/j.1540-5907.2006.00218.x.

Staiger, D., & Stock, J. (1997). Instrumental variables regression with weak instruments. *Econometrica*, *65*(3), 557-586. https://doi.org/10.2307/2171753.

Waldron, I. (1991). Patterns and causes of gender differences in smoking. *Social Science & Medicine*, *32*(9), 989-1005. https://doi.org/10.1016/0277-9536(91)90157-8.

WHO (2020). Smoking and COVID-19. Scientific Brief. Online: WHO/2019-nCoV/Sci_Brief/Smoking/2020.1.

Zhang L, Ghader S, Pack M, Darzi A, Xiong C, Yang M, Sun Q, Kabiri A, Hu S. (2020). An interactive COVID-19 mobility impact and social distancing analysis platform. medRxiv 2020. https://doi.org/10.1101/2020.04.29.20085472.

Appendix A

Table A1: Description of the variables and summary statistics

Variable name	Variable description	Year	Mean	St. Dev.	Source
Covid_Cases	Number of registered cases of COVID-19 per million population in logarithms (as of June 8, 2020)	2020	8.38	0.76	[1]
Covid_CFR	Case fatality rate, % (as of June 8, 2020)	2020	4.3	1.97	[1]
Excess_Deaths	Number of estimated excess deaths in the period between February 1 and June 3 cases per million population plus one, in logarithms	2020	3.58	2.16	[2]
Mortality	Number of deaths per million population, in logarithms	2019	9.09	0.14	[3]
Smoking	Share of population who smoke, %	2018	16.61	3.33	[4]
Smoking_Daily	Share of population who smoke daily, %	2018	11.73	2.99	[4]
Cigarettes	The logarithm of packs of cigarettes purchased over the year per capita	2018	3.63	0.47	[5]
Cigarette_Tax	State tax for the pack of cigarettes, in dollars	2018	1.7	1.11	[5]
Neighbor_Tax	Average state-taxes for the pack of cigarettes in the neighboring states in 2013, in dollars	2013	1.39	0.72	[5]
Population	Logarithm of population	2019	15.23	1.03	[3]
Population_65	Share of population older than 65	2018	16.46	1.87	[3]
Population_45	Share of population in the age between 45 and 64	2018	25.65	1.8	[3]
Density	The logarithm of people per square mile	2020	4.72	1.43	[3]
White	An estimated share of the white population, %	2017	69.21	15.1	[3]
Income	Average annual income adjusted for the cost of living	2019	10.88	0.11	[6]
Inequality	Gini index, average over 2013-2017 years	2017	46.63	1.99	[3]
Poverty	Percentage of population living below the federal poverty line, average over 2013-2017 years	2017	0.14	0.03	[3]
Healthiness	Share of the surveyed population who report an excellent state of health	2018	17.62	2.36	[4]
Overweight	Share of the surveyed population who report their BMI being over 25	2018	66.31	3.37	[4]
Stayhome	Percentage of residents staying at home for the period March 1 – May 31	2020	0.28	0.05	[7]
Gov_Republican	A dummy that equals one if the governor of the state is a Republican	2020	0.51	0.51	[8]

Sources: [1] Worldometer.info
[2] CDC
[3] US Census Bureau
[4] CDC: BRFSS
[5] CDC: State Tobacco Activities Tracking and Evaluation
[6] Bureau of Economic Analysis
[7] Maryland Transportation Institute
[8] Ballotpedia

Table A.2: The effect of smoking prevalence on the severity of COVID-19 with controls

	(1)	(2)	(3)
Dep. Variable:	Covid_Cases	Covid_CF	Excess_Deaths
Smoking	-0.301***	-1.001***	-0.842**
	(-3.27)	(-2.83)	(-2.13)
Population	-0.354***	-0.498	-0.507
	(-3.65)	(-1.25)	(-1.14)
Population_65	-0.112**	-0.247	-0.257
	(-2.01)	(-1.56)	(-1.41)
Population_45	0.11	0.944***	0.831***
	(1.56)	(3.92)	(3.11)
Density	0.236***	0.113	0.505
	(2.83)	(0.31)	(1.52)
Income	3.366***	9.799***	11.014***
	(4.84)	(3.20)	(3.36)
Inequality	0.062	0.04	-0.456
	(0.86)	(0.11)	(-1.04)
Poverty	0.148*	0.962**	0.949**
	(1.74)	(2.57)	(2.37)
Healthiness	-0.143*	-0.334	-0.245
	(-1.89)	(-0.97)	(-0.65)
Overweight	0.150**	0.231	0.224
	(2.23)	(0.99)	(0.85)
White	0.021*	0.127***	0.053
	(1.69)	(2.62)	(1.07)
Gov_Republican	-0.048	-0.036	0.737
	(-0.28)	(-0.07)	(1.24)
Stayhome	0.082***	0.336**	0.349**
	(2.64)	(2.43)	(2.48)
R2	0.588	0.248	0.36

Notes: t statistics in parentheses, * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$. Number of states N=49.

Table A.3: Robustness check, exclusion of states of New York and New Jersey

	(1)	(2)	(3)
Dep. Variable:	Covid_Cases	Covid_CFR	Excess_Deaths
Smoking	-0.272***	-1.137***	-0.912**
	(-2.71)	(-3.04)	(-2.03)
Controls	Yes	Yes	Yes
F-statistics	16.69	16.69	16.69
LM test	0.001	0.001	0.001

Notes: t statistics in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01. Number of states N=47.

Table A.4: Robustness check, controlling for the neighboring spread of COVID-19 cases and excess mortality

	(1)	(2)	(3)
Dep. Variable:	Covid_Cases	Covid_CFR	Excess_Deaths
Smoking	-0.216***	-0.999***	-0.721*
	(-2.79)	(-2.64)	(-1.79)
Neighbor_Cases	0.650**	-0.192	1.17
	(2.05)	(-0.14)	(1.02)
Neighbor_Excess Deaths	-0.122	0.085	-0.277
	(-1.01)	(0.18)	(-0.68)
Controls	Yes	Yes	Yes
F-statistics	12.62	12.62	12.62
LM test	0.004	0.004	0.004

Notes: t statistics in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01. Number of states N=49.

Table A.5: Robustness check, using different year for neighboring cigarette-taxes (2010)

	(1)	(2)	(3)
Dep. Variable:	Covid_Cases	Covid_CFR	Excess_Deaths
Smoking	-0.302***	-0.935***	-0.898**
	(-3.56)	(-2.84)	(-2.26)
Controls	Yes	Yes	Yes
F-statistics	18.29	18.29	18.29
LM test	0.001	0.001	0.001

Notes: t statistics in parentheses, * p < 0.1, ** p < 0.05, *** p < 0.01. Number of states N=49.