

---

# COVID-19 disease Severity Correlates with Smoking Status


---

 Édouard Lansiaux

Henry Warembourg School of Medicine, Medical Student

Lille University, 59000, Lille, France

[edouard.lansiaux.etu@univ-lille.fr](mailto:edouard.lansiaux.etu@univ-lille.fr)

 Philippe P. Pébaÿ Ph.D.

NexGen Analytics

Sheridan, WY 82801, U.S.A.

[philippe.pebay@ng-analytics.com](mailto:philippe.pebay@ng-analytics.com)

Jean-Laurent Picard, M.Sc

Conservatoire National des Arts et Métiers

75141 Paris, France

[jl@pi.cards](mailto:jl@pi.cards)

Joachim Son-Forget, M.D.-Ph.D.

Assemblée Nationale

75355, Paris, France

[joachim.son-forget@assemblee-nationale.fr](mailto:joachim.son-forget@assemblee-nationale.fr)

April 23<sup>th</sup>, 2020

All authors have completed and submitted the International Committee of Medical Journal Editors form for disclosure of potential conflicts of interest. No potential conflicts of interest were disclosed.

## ABSTRACT

The novel COVID-19 disease is a contagious acute respiratory infectious disease whose causative agent has been demonstrated to be a new virus of the coronavirus family, SARS-CoV-2. Multiple studies have already reported that risk factors for severe disease include older age and the presence of at least one of several underlying health conditions. However, a recent physiopathological report and the French COVID-19 scientific council have postulated a protective effect of tobacco smoking.

We have been able to demonstrate the statistical significance in this regard of recent series from both China and in the US, reporting smoking status as well as disease severity (p-values of  $2.27 \cdot 10^{-3}$  and  $11.7 \cdot 10^{-15}$ , respectively). Subsequently and using a Bayesian approach we have established that disease severity is positively associated with smoking status. Finally, we refute claims linking general population smoking status ( $N=O(10^8)$  or  $O(10^9)$ ) to much smaller disease course series ( $N=O(10^3)$ ). The latter point in particular is presented to stimulate academic discussion, and must be further investigated by well-designed studies.

Keywords: COVID-19; Coronavirus; Respiratory Distress; Tobacco; Smoking; Correlation; Statistics; Conditional; Probability; Regression; China; U.S.A.

## IMPLICATIONS

Recent preliminary estimates of the prevalence of selected underlying health conditions among patients with COVID-19, in China as well as in the United States, show that individuals with underlying health conditions appear to be at higher risk for more severe COVID-19. Thanks to a non parametric test, we have concluded that the only significant findings that can be confidently asserted from the series discussed in this article, is that COVID-19 disease severity is positively associated with smoking status.

## INTRODUCTION

The novel coronavirus pneumonia (COVID-19) is a contagious acute respiratory infectious disease whose causative agent has been demonstrated to be a new virus of the coronavirus family, SARS-CoV-2. This illness was first evinced in December 2019 in the Seafood Market of Wuhan, Hubei Province, in southern China<sup>1,2</sup>. Patients with the coronavirus pneumonia have a fever, and the temperature above 38 degrees Celsius with symptoms such as dry cough, fatigue, dyspnea, difficulty breathing, and diarrhea<sup>1-5</sup>. There is a high transmission for this disease, which, as result of this and other factors such as international travel and trade, has now turned into a pandemic, with hundred thousands of individuals confirmed to be infected worldwide – and most likely millions of unreported cases<sup>5</sup>.

Recent preliminary estimates of the prevalence of selected underlying health conditions among patients with COVID-19, in China as well as in the United States, show that individuals with underlying health conditions appear to be at higher risk for more severe COVID-19<sup>3,6</sup>. A recent Greek epidemiological review suggests that smoking is most likely associated with the negative progression and adverse outcomes of COVID-19<sup>7</sup>. Meanwhile, a recent physio-pathology study has formulated the hypothesis that the effectiveness of ACE inhibitors/angiotensin II receptor blockers and nicotine in patients infected with COVID-19<sup>8</sup>.

The smoking prevalence in China and in the U.S. have been previously and broadly documented: these are, respectively, of 26.6%, and 19.7%<sup>9,10</sup>. Furthermore, the president of the COVID-19 Scientific Council of France recently postulated that nicotine may have a protective role against COVID-19 infection<sup>11</sup>. Thanks to these multiple studies, we can observe that the prevalence of past/current smokers is higher in COVID-19 severe infected patients. In this context, we would like to know whether the past/current smoking status entails higher probability of a severe course of COVID-19 illness, as compared to having never smoked.

# METHODS

## Study and Participants

We conducted a descriptive observational cross-sectional study in order to define a hypothetical relationship between smoker status and SARS-CoV-2 severity infection. The source and targeted populations are the whole humanity in view of the ongoing COVID-19 pandemic. The eligible population is constituted by all Chinese and American already included in previous papers.

The study was conducted by a consortium of two data analysts, a MD specialized in radiology, and a medical student in clinical years. NexGen Analytics had no role in designing the study, nor in making the decision to submit manuscript to the publication, nor did it receive any fee or compensation in the context of this work. The first author vouches for the data and analyses, as well as for the fidelity of this report to the study protocol.

## Enrollment

The studied and included population gathers patients from two studies: one group of patients studied in China between 2019/12/11 and 2020/01/31 ( $N = 1,085$ ), and another group studied in the U.S. between 2020/02/12 and 2020/03/28 ( $N = 6,637$ )<sup>3,6</sup>.

## Outcome Measures

Our null hypothesis ( $H_0$ ) was the orthogonality between two variables: the smoking status ( $X$ ) and the COVID-19 Symptoms Severity ( $Y$ ). Each variable can take two values: for the smoking status ( $X$ ), it was a never smoker ( $X_1$ ) or a current/past smoker ( $X_2$ ); for the COVID-19 Symptoms Severity ( $Y$ ), it was a mild (non-hospitalised) ( $Y_1$ ) or a severe (hospitalised, ICU or not) ( $Y_2$ ) symptoms severity.

## Statistical Analysis

Each sample was analysed separately. First, we had to calculate for each sample different conditional probabilities: the probability to develop severe symptoms for non-smokers (i.e.,  $P(Y_2|X_1)$ ) and the probability to develop severe symptoms for former and current smokers combined (i.e.,  $P(Y_2|X_2)$ ). We then calculated, for each observed sample, the expected cardinalities assuming  $H_0$  expected sample: in each case, we multiplied the two corresponding sums and subsequently divided by the actual total. In this way, we were able to apply a  $\chi^2$  test to compare the observed table to the expected table, for each sample. The p-value was subsequently computed in order to evince a statistically significant difference between the observed and the expected tables.

## RESULTS

### Populations

After a short treatment of the initial data from both Chinese and U.S. series (*Table 1*), in particular to eliminate incomplete data, we have obtained the observed data for each sample with their respective conditional probabilities (*Tables 2, 3*). For the Chinese sample, the probability to develop severe COVID-19 symptoms for non-smokers is of 14.5% and the probability to develop severe COVID-19 symptoms for former/current smokers is of 24.1%. For the American sample, the probability to develop severe COVID-19 symptoms for non-smokers is of 21.7% and the probability to develop severe COVID-19 symptoms for former/current smokers is of 42.7%.

### Outcomes

We estimated the expected data on each case by multiplying the two corresponding sums and then by dividing by the total effective, for each sample (*Tables 4, 5*). In these conditions, we were able to estimate the  $\chi^2$  parameter, for Chinese and American samples, and the p-values where the difference

between observed and expected data are significant (*Table 6*). Indeed, the difference observed between them is significant for China and US with respectively a p-value of  $2.27 \cdot 10^{-3}$  and  $11.7 \cdot 10^{-15}$ .

## DISCUSSION

Subsequently, by computing the empirical probability distributions conditioned on smoking status, we have established that disease severity is positively associated with smoking status, in those samples. Nevertheless, our study has some limitations:

1. The difference between smoking status as observed across the entire Chinese and American populations<sup>9,10</sup>, and the same observed in our small clinical datasets (*Table 1*) is significant. As a result, the previously claimed relationship between former/current smoking status and a COVID-19 disease severity may not be confidently estimated from a global population.
2. Only adults were considered to compute smoking prevalence<sup>10</sup>.

However, regarding the first limitation above, we contend that claims linking general population smoking status, computed across populations with cardinalities  $N$  in the order of  $O(10^8)$  or even  $O(10^9)$  with broadly varying smoking habits (e.g. already rural vs urban divide in this regard), to much smaller disease course series ( $N = O(10^3)$ ), are at best imprudent and at worst marred by selection bias, especially as the latter regard a fraction of patients from the same urban area. We therefore claim that specific smoking status data, relevant to the concerned geographic areas, is required to evince a statistically significant under-representation of smokers among COVID-19 patients. An alternative argument could be made from a physio-pathological standpoint as well: indeed, previous studies and assertions describe a hypothetical protector role of nicotine against COVID-19 infection<sup>8</sup>, but these are largely speculative at this point and require laboratory confirmation. We therefore conclude that the only significant findings that can be confidently asserted from the series discussed in this article, is that COVID-19 disease severity is positively associated with smoking status.

## **ACKNOWLEDGEMENTS**

The authors declare no conflicts of interest.



## REFERENCES

1. Wang D, Hu B, Hu C, et al. Clinical Characteristics of 138 Hospitalized Patients With 2019 Novel Coronavirus–Infected Pneumonia in Wuhan, China. JAMA. 2020;323(11):1061–9.
2. Huang C, Wang Y, Li X, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet. 2020 Jan 24;395.
3. Guan W-J, Ni Z, Hu Y, et al. Clinical Characteristics of Coronavirus Disease 2019 in China. New England Journal of Medicine. 2020 Feb 28;
4. Chang D, Lin M, Wei L, et al. Epidemiologic and Clinical Characteristics of Novel Coronavirus Infections Involving 13 Patients Outside Wuhan, China. JAMA. 2020 Feb 7;323.
5. Diao K, Han P, Pang T, Li Y, Yang Z. HRCT Imaging Features in Representative Imported Cases of 2019 Novel Coronavirus Pneumonia. Precision Clinical Medicine. 2020 Feb 11;
6. Chow N, Fleming-Dutra K, Gierke R, et al. Preliminary Estimates of the Prevalence of Selected Underlying Health Conditions Among Patients with Coronavirus Disease 2019 — United States, February 12–March 28, 2020. Morbidity and Mortality Weekly Report. 2020 Apr 3;69(13).

7. Vardavas C, Nikitara K. COVID-19 and smoking: A systematic review of the evidence. Tobacco Induced Diseases. 2020 Mar 20;18.

8. Patel A, Verma A. COVID-19 and Angiotensin-Converting Enzyme Inhibitors and Angiotensin Receptor Blockers: What Is the Evidence? JAMA. 2020 Mar 24;

9. World Health Organization, US Center for Disease Control and Prevention, RTI International. Western Pacific Region - China - National Population Survey - Global Adult Tobacco Survey (GATS) Topic: Tobacco Use [Internet]. Global Tobacco Surveillance System Data (GTSSData). Available from:[https://nccd.cdc.gov/GTSS/rdPage.aspx?rdReport=OSH\\_GTSS.ExploreByLocation&isCountry=ch&isDataSource=GATS&isLocation=1475&isTopic=T06&isWHORegion=WPR&isYear=20102010&rdRnd=93116](https://nccd.cdc.gov/GTSS/rdPage.aspx?rdReport=OSH_GTSS.ExploreByLocation&isCountry=ch&isDataSource=GATS&isLocation=1475&isTopic=T06&isWHORegion=WPR&isYear=20102010&rdRnd=93116)

10. Creamer M, Wang T, Babb S, et al. Tobacco Product Use and Cessation Indicators Among Adults - United States, 2018. MMWR Morbidity and mortality weekly report. 2019 Nov 15;68:1013–9.

11. Deszpot T. Le tabac, une barrière contre le Covid-19 ? Attention aux fausses infos [Internet]. LCI. 2020. Available from: <https://www.lci.fr/sante/coronavirus-le-tabac-protege-du-covid-19-attention-aux-fausses-infos-sur-les-fumeurs-2150848.html>

## TABLES

Sample	China			United States			
Sub-sample	All patients	Non severe	Severe	All patients	Not hospitalized	Hospitalized	ICU
<b>Cardinality</b>	1,085	913	172	6,637	5,143	1,037	457
<b>Never smoke (%)</b>	927 (85.4)	793 (86.9)	134 (77.9)	6,391 (96.3)	5,002 (97.2)	970 (93.5)	419 (91.7)
<b>Former smoker (%)</b>	21 (1.9)	12 (1.3)	9 (5.2)	132 (2.0)	80 (1.6)	45 (4.3)	33 (7.2)
<b>Current smoker (%)</b>	137 (12.6)	108 (11.8)	29 (16.9)	88 (1.7)	61 (1.2)	22 (2.2)	5 (1.1)

*Table 1. Reported outcomes concerning smoking status on two samples.*

China observed	$Y_1$	$Y_2$	TOTAL
$X_1$	793	134	927
$X_2$	120	38	158
TOTAL	913	172	1,085

*Table 2. Reported outcomes from the China sample.*

X : smoking status

1.  $X_1$ : never smoke

2.  $X_2$ : current or former smoker

Y : severity of symptoms

1.  $Y_1$ : mild (non-hospitalized)

2.  $Y_2$ : severe (hospitalized, in the ICU or not)

U.S. observed	$Y_1$	$Y_2$	TOTAL
$X_1$	5,002	1,389	6,391
$X_2$	141	105	246
TOTAL	5,143	1,454	6,637

*Table 3. Reported outcomes from the U.S. sample.*

X : smoking status

1.  $X_1$ : never smoke

2.  $X_2$ : current or former smoker

Y : severity of symptoms

1.  $Y_1$ : mild (non-hospitalized)

2.  $Y_2$ : severe (hospitalized, in the ICU or not)

China expected	$Y_1$	$Y_2$	TOTAL
$X_1$	780	147	927
$X_2$	133	25	158
TOTAL	913	172	1,085

*Table 4. Expected outcomes from the China sample.*

X : smoking status

1.  $X_1$ : never smoke

2.  $X_2$ : current or former smoker

Y : severity of symptoms

1.  $Y_1$ : mild (non-hospitalized)

2.  $Y_2$ : severe (hospitalized, in the ICU or not)

<b>U.S. expected</b>	<b>Y<sub>1</sub></b>	<b>Y<sub>2</sub></b>	<b>TOTAL</b>
<b>X<sub>1</sub></b>	4,952	1,459	6,391
<b>X<sub>2</sub></b>	191	55	246
<b>TOTAL</b>	5,143	1,454	6,637

*Table 5. Expected outcomes from the U.S. sample.*

X : smoking status

1. X<sub>1</sub>: never smoke

2. X<sub>2</sub>: current or former smoker

Y : severity of symptoms

1. Y<sub>1</sub>: mild (non-hospitalized)

2. Y<sub>2</sub>: severe (hospitalized, in the ICU or not)

Sample	China	U.S.
$\chi^2$	9.32	59.6
<b><i>p</i>-value</b>	$2.27 \cdot 10^{-3}$	$11.7 \cdot 10^{-5}$

*Table 6.  $\chi^2$  test applied to both samples.*