

Letter to the Editor

# Minimizing Selection and Classification Biases. Comment on “Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing”

Francisco Martos Pérez<sup>1</sup>, MD, PhD; Ricardo Gomez Huelgas<sup>2</sup>, MD, PhD; María Dolores Martín Escalante<sup>1</sup>, MD, PhD; José Manuel Casas Rojo<sup>3</sup>, MD

<sup>1</sup>Department of Internal Medicine, Hospital Costa del Sol, Marbella, Spain

<sup>2</sup>Internal Medicine Department, Hospital Regional de Málaga, Málaga, Spain

<sup>3</sup>Internal Medicine Department, Infanta Cristina University Hospital, Parla (Madrid), Spain

**Corresponding Author:**

Francisco Martos Pérez, MD, PhD

Department of Internal Medicine

Hospital Costa del Sol

Autovía A-7, Km 187

Marbella, 29603

Spain

Phone: 34 658927715

Email: [pacomartos1@gmail.com](mailto:pacomartos1@gmail.com)

**Related Articles:**

Comment on: <https://www.jmir.org/2020/10/e21801/>

Comment in: <https://www.jmir.org/2021/5/e29405/>

(*J Med Internet Res* 2021;23(5):e27142) doi: [10.2196/27142](https://doi.org/10.2196/27142)

**KEYWORDS**

artificial intelligence; big data; COVID-19; electronic health records; tachypnea; SARS-CoV-2; predictive model; prognosis; classification bias; critical care

The paper by Izquierdo et al [1], published in the recent issue of the *Journal of Medical Internet Research*, employed a combination of conventional and machine learning tools to describe the clinical characteristics of patients with COVID-19 and the factors that predict intensive care unit (ICU) admission. We would like to make some comments about its design.

The authors should have provided the proportion of patients with a positive microbiological diagnosis. If the artificial intelligence software's capacity for retrieving this information is limited in some way, this should be explained. The classification bias introduced by the lack of microbiological confirmation may have been significant since the study includes patients from January 1, 2020. Although some undiagnosed cases have likely been present prior to the first declared case (March 1, 2020) [2] in Castilla-La Mancha, it is improbable that there were many of them.

ICU admissions are related to many factors not addressed in the study. The decision not to admit a patient to the ICU because of short life expectancy, low quality of life, or high burden of comorbidities may have had a great impact during the first wave

of the COVID-19 pandemic, when a scarcity of ICU beds manifested in some regions of Spain. The 6.1% ICU admission rate reported by the authors was 36% lower than the 8.3% reported in a national survey of 15,111 patients from 150 hospitals in Spain [3]. We could hypothesize that the patients included in the study had a milder form of the disease. However, given the absence of a microbiological diagnosis in an unknown percentage of patients, the inclusion of a significant proportion of patients without a real COVID-19 diagnosis cannot be ruled out. These doubts could have been resolved if a microbiological diagnosis had been a requisite for inclusion. The mortality rate, the most robust and relevant endpoint, should also have been reported and the factors related to it analyzed.

Artificial intelligence is having an increasing impact on the rate of health care information processing. However, minimization of selection and classification biases should be guaranteed in the design of investigations. In this case, this could have been achieved by including only microbiologically confirmed cases and prolonging the period of inclusion, since most COVID-19 cases emerged after the end date of the study inclusion period.

These changes in the design would have allowed for a better evaluation of the performance of artificial intelligence techniques, making the results obtained in the sample closer to those of the real population.

### Conflicts of Interest

None declared.

### References

1. Izquierdo JL, Ancochea J, Savana COVID-19 Research Group, Soriano JB. Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing. *J Med Internet Res* 2020 Oct 28;22(10):e21801 [FREE Full text] [doi: [10.2196/21801](https://doi.org/10.2196/21801)] [Medline: [33090964](https://pubmed.ncbi.nlm.nih.gov/33090964/)]
2. Europa Press. 2020 Mar 1. URL: <https://www.europapress.es/castilla-lamancha/noticia-varon-62-anos-ingresado-guadalajara-primer-caso-positivo-coronavirus-lm-20200301103741.html> [accessed 2020-01-08]
3. Casas-Rojo JM, Antón-Santos JM, Millán-Núñez-Cortés J, Lumbreras-Bermejo C, Ramos-Rincón JM, Roy-Vallejo E, et al. Clinical characteristics of patients hospitalized with COVID-19 in Spain: Results from the SEMI-COVID-19 Registry. *Rev Clin Esp (Barc)* 2020 Nov;220(8):480-494 [FREE Full text] [doi: [10.1016/j.rce.2020.07.003](https://doi.org/10.1016/j.rce.2020.07.003)] [Medline: [32762922](https://pubmed.ncbi.nlm.nih.gov/32762922/)]

### Abbreviations

**ICU:** intensive care unit

*Edited by T Derrick; submitted 12.01.21; peer-reviewed by M Elbattah, J Marconato, T Capote; accepted 13.05.21; published 26.05.21*

*Please cite as:*

*Martos Pérez F, Gomez Huelgas R, Martín Escalante MD, Casas Rojo JM*

*Minimizing Selection and Classification Biases. Comment on “Clinical Characteristics and Prognostic Factors for Intensive Care Unit Admission of Patients With COVID-19: Retrospective Study Using Machine Learning and Natural Language Processing”*  
*J Med Internet Res* 2021;23(5):e27142

URL: <https://www.jmir.org/2021/5/e27142>

doi: [10.2196/27142](https://doi.org/10.2196/27142)

PMID: [33989163](https://pubmed.ncbi.nlm.nih.gov/33989163/)

©Francisco Martos Pérez, Ricardo Gomez Huelgas, María Dolores Martín Escalante, José Manuel Casas Rojo. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 26.05.2021. This is an open-access article distributed under the terms of the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work, first published in the Journal of Medical Internet Research, is properly cited. The complete bibliographic information, a link to the original publication on <https://www.jmir.org/>, as well as this copyright and license information must be included.