Be careful with Big Data: Re-analysis of Patient Characteristics and Outcomes of 11,721 Patients with COVID19 Hospitalized Across the United States

ROUSSEL Yanis[1,2], MILLION Matthieu[1,2], CHABRIERE Eric[1,2], LAGIER Jean-Christophe[1,2], DIDIER Raoult[1,2]

Affiliations :

[1] IHU-Méditerranée Infection, Marseille, France

[2] Aix Marseille Univ, IRD, AP-HM, MEPHI, Marseille, France

Corresponding author: Didier Raoult, IHU - Méditerranée Infection, 19-21 boulevard Jean Moulin, 13005 Marseille, France. Tel.: +33 413 732 401, Fax: +33 413 732 402; email: didier.raoult@gmail.com

Dear editor,

We read with interest the article published by Fried *et al.* (1) in Clinical Infectious Diseases. We wish to alert the editor about the robustness of this work. First of all, the groups are not homogeneous, such as in other COVID-19 big-data studies (2). More patients were intubated in the hydroxychloroquine group than in the non-hydroxychloroquine group (24.9% vs 12.2%). Seven percent of patients were intubated on the first day of hospitalization. While the mechanical ventilation subgroup of patients in the remdesivir group is analyzed in the supplementary data, this is not the case for the hydroxychloroquine group. 93% of patients in the hydroxychloroquine group had pneumonia, versus 79% in the non-hydroxychloroquine group (Supplemental table 3 in Field and al (1)). The risk of being trapped in a Simpson's paradox-like situation (3) is high.

This brings us to our main point of concern, which is the reliability of the data used in this article. Among the 11 authors, 7 are affiliated to a data collection company, namely Target Pharmasolutions. In this article, they provide little detail on how the data were collected. The authors state that the data come "from a commercial insurance claims database that requires a data sharing agreement and data license for access". They also specify that the data "were acquired from a commercially available source representing adults receiving inpatient care between February 15 and April 20, 2020 at 245 hospitals across 38 states in the US". The hospital names are not provided, nor whether these hospitals have agreed to have their data used in such a study. The information available on the Target Pharmasolutions company website does not provide further details on the data collection mechanism.

There are some points that catch our attention. For instance, we do not understand how 99.4% of patients treated with hydroxychloroquine were treated in urban hospitals, compared to 65% of untreated patients (Supplemental Table 3 in Fried *et al*.(1)), while patients are distributed in a more balanced manner between teaching or not-teaching hospitals, as well as in the most urbanized (Northeast) and less urbanized (Midwest) regions of the United States. Likewise, the mortality rate of 70.5% among patients under mechanical ventilation (Table 2 in Fried *et al*.(1)) does not appear to us to be compatible with the data published in the literature. No information is given on any verification of individual data by any infectious diseases specialist, so it is possible that patient registration errors have been included in the general study.

The scandal caused by the retraction of two articles initially published in The Lancet (4) and in the New England Journal of Medicine (5) by a data collection company, Surgisphere, which was unable to demonstrate the reliability of its data, emphasized the importance of data traceability. We therefore ask the editors of Clinical Infectious Diseases to ensure the scientific community, as the Lancet editors did for the article published by Merah et al, that the data presented in the article published by Fried *et al*. meet the most essential reliability criteria.

Acknowledgements

None

Funding

None

Declaration of competing interest

The authors declare no competing interests. Funding sources had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; and preparation, review, or approval of the manuscript. Our Marseille group used widely available generic drugs distributed by many pharmaceutical companies.

References

(1) Fried MW, Crawford JM, Mospan AR, Watkins SE, Munoz Hernandez B, Zink RC, Elliott S, Burleson K, Landis C, Reddy KR, Brown RS. Patient Characteristics and Outcomes of 11,721 Patients with COVID19 Hospitalized Across the United States. Clin Infect Dis. 2020 Aug 28:ciaa1268. Doi: 10.1093/cid/ciaa1268. Epub ahead of print. PMID: 32856034; PMCID: PMC7499515.

(2) R. Clinton Ohlers, Effectiveness of hydroxychloroquine was hiding in plain sight, World Tribune, published online September 21, 2020.

(3) Simpson EH. The Interpretation of Interaction in contingency tables. Journal of the Royal Statistical Society, série B, vol. 13, 1951, p. 238-241. https://doi.org/10.1111/j.2517-6161.1951.tb00088.x

(4) Mehra MR, Desai SS, Ruschitzka F, Patel AN. RETRACTED: Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. Lancet. 2020 May 22:S0140-6736(20)31180-6. doi: 10.1016/S0140-6736(20)31180-6. Epub ahead of print. Retraction in: Lancet. 2020 Jun 5;:null. Erratum in: Lancet. 2020 May 30;: PMID: 32450107; PMCID: PMC7255293.

(5) Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. N Engl J Med. 2020 Jun 18;382(25):e102. doi: 10.1056/NEJMoa2007621. Epub 2020 May 1. Retraction in: N Engl J Med. 2020 Jun 4;: PMID: 32356626; PMCID: PMC7206931.