The first 10,000 COVID-19 papers in perspective: Are we publishing what we should be publishing?

Anna Odone, MD PhD
Associate Professor of Public Health
School of Medicine, Vita-Salute San Raffaele University
Milan (Italy)

Carlo Signorelli, MD PhD
Professor of Public Health
School of Medicine, Vita-Salute San Raffaele University
Milan (Italy)

David Stuckler, MPH PhD
Professor of Political Economy and Sociology
Department of Social and Political Sciences,
Bocconi University
Milan (Italy)

Sandro Galea, MD, DrPH
Dean
Robert A Knox Professor
School of Public Health, Boston University
Boston, MA (USA)

and the **University Vita-Salute San Raffaele COVID-19 literature monitoring working group:** Andrea Amerio MD PhD, Lorenzo Bellini MD, Daria Bucci, Michele Capraro MD, Giovanni Gaetti, MD Stefano Salvati MD

Corresponding author:

Prof. Anna Odone
School of Medicine, Vita-Salute San Raffaele University
IRCCS San Raffaele Scientific Institute
Via Olgettina 60, 20132 Milan (Italy)
Tel: +39.02.26435525
Mail: anna.odone@mail.harvard.edu

The Covid-19 pandemic has led to an unprecedented focus of the world's scientific community on one topic. To quantify, we have calculated that 4% of all scientific outputs during the last five months have been about COVID-19; this has increased from 0.3% in February, to 1.2% in March, 4.5% in April, 6.5% in May, 8.3% in June and 6.6% in July. We systematically retrieved and critically assessed the first 10,000 Pubmed indexed papers on COVID-19. They were published between 20th January and 7th May 2020, with an average of nearly 100 new papers added every day, published in 1881 different scientific journals. Fewer than 8% of journals have published half of the total production, and 7 journals alone have indexed more than 100 papers each. By contrast, 43.3% of journals only published one paper on COVID-19. Unsurprisingly, the largest amount of papers, one-fourth of the 10,000, were published in the US, the country with the largest COVID-19 burden and ranking first in the 2019 *Nature index* for quality research[1], followed by China (22.2%), Italy (9%), the UK (7.6%) and France (3.2%).

Are we publishing what we should be publishing? At the end of January, *Nature* listed the six key questions scientists should be asking[2]. Among these were: how does the virus spread? Can infected people spread the virus without showing symptoms? How deadly is the virus? Where did the virus come from? What can we learn from the virus's genetic sequence? Can a drug be developed to treat the coronavirus? Although acknowledging that some research questions are quicker to answer than others, and that some research answers are moving targets, almost five months after none of them is fully answered. Later in February 2020, a more detailed list of research priorities was identified as essential to inform effective public health responses to COVID-19[3]. In March, the *Science Translational Medicine* Editorial framed key questions for pandemic prevention, identifying selected pathogen- and society-based variables to be measured[4]. We are far from answering these questions, from fully understanding COVID-19 epidemiology, or from having quantified the impact of different control measures. While science takes time, and not understanding is not an indictment of purposeful science, it is worth asking whether the deluge of science has been asking the right questions. The largest share of COVID-19 papers thus far have focused on clinical management descriptions of hospitalized cases, and reflections on the implications of the COVID-19 emergency on different clinical specialties (29.7%). Over time, the percentage of papers reporting surveillance or epidemiological data have been decreasing (from 56% of all COVID-19 papers at the beginning of February, to 10% at the beginning of May); little has so far been published on new therapies and treatment evaluation (4.4%), although trends in this regard are increasing. Other COVID-19-related papers in the literature include health services research (6.3%), mental health (3.5%), aspects related to communication in times of emergency (2.5%), and economic impacts (0.5%).

Are we publishing the way we should be publishing? As scientific output around COVID-19 evolves over time, we find, consistent with other efforts that are systematically monitoring the literature[5], both poor adherence to identified research priorities, and a predominance of opinion over data. More than 60% of published papers on COVID-19 are opinion pieces not reporting original data.

Overall, the question remains: has the surge in scientific publication around COVID-19 been a positive, or a negative, for science? On the positive side, the COVID-19 public health emergency context has pushed journals to laudable efforts to fast track peer reviews, publishers to waive publication fees and provide free access to articles' content, and encouraged a preprint model of publication, the latter carrying both pro and con arguments. It has highlighted the role of new tools based on machine learning and artificial intelligence that are available to support methodologists in conducting systematic reviews or assessing research quality[6]. And it has shown us that scientific publications might become live documents, constantly updated. On the negative side, this moment has made foggier the distinction between data-driven and expository outputs, with important implications for how the work of science is communicated. It also has revealed a divide between the production of science with its ultimate aims, of taking us towards individual and population wellbeing.

Galileo Galilei in the *Dialogue Concerning the Two Chief World Systems*[7] warned science had to deal with the "sensible world and not a world of paper". How would he react to so few published papers on COVID-19 report original data? How would he react to the rapid dissemination of inaccurate and exaggerated information?[8] The idea of a "sensible world" was the revolt of scientists against philosophers writing their opinions devoid of empiric observation and physical fact. One wonders whether we are entering a new Galilean age where science and empiricism needs to regain the upper hand, focusing on questions that address key scientific need, and prioritizing data over opinion in an effort to solve a global problem.

References

1. Nature. Nature Index. Available at: https://www.natureindex.com/annual-tables/2020/country/all [Accessed: 17.05.2020].

2. Callaway E, Cyranoski D. China coronavirus: Six questions scientists are asking. *Nature* 2020; **577**(7792): 605-7.

3. Cowling BJ, Leung GM. Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. *Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin* 2020; **25**(6).

4. Layne SP, Hyman JM, Morens DM, Taubenberger JK. New coronavirus outbreak: Framing questions for pandemic prevention. *Science translational medicine* 2020; **12**(534).

5. National Institute for Health Research Policy Research Programme Reviews. COVID-19: living map of the evidence. Available: http://eppi.ioe.ac.uk/COVID19_MAP/covid_map_v7.html [Accessed: 17.05.2020]

6. Khamsi R. Coronavirus in context: Scite.ai tracks positive and negative citations for COVID-19 literature. *Nature* 2020.

7. Galileo Galilei. Dialogue Concerning the Two Chief World Systems (Dialogo sopra i due massimi sistemi del mondo). First edition: Florence, 1632

8. Ioannidis JPA. Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. *European journal of clinical investigation* 2020; **50**(4): e13222.