

The runaway science: a bibliometric analysis of the COVID-19 scientific literature

Anna Odone^{1,2}, Stefano Salvati¹, Lorenzo Bellini¹, Daria Bucci¹, Michele Capraro¹, Giovanni Gaetti¹, Andrea Amerio^{3,4,5}, Carlo Signorelli¹

¹School of Medicine, University Vita-Salute San Raffaele, Milan, Italy; ²Clinical Epidemiology and HTA, IRCCS San Raffaele Scientific Institute, Milan, Italy; ³Department of Neuroscience, Rehabilitation, Ophthalmology, Genetics, Maternal and Child Health (DINOGMI), Section of Psychiatry, University of Genoa, Genoa, Italy; ⁴IRCCS Ospedale Policlinico San Martino, Genoa, Italy; ⁵Department of Psychiatry, Tufts University, Boston, MA, USA

Abstract. *Background and aim of the work:* To reflect on content, trends and quality of scientific publishing on COVID-19. In particular, to report on the systematic screening, quantitative assessment and critical appraisal of the first 10,000 scientific papers published on COVID-19 and to compare how scientific outputs matched identified research priorities and public health needs. *Methods:* A comprehensive research strategy was developed to systematically retrieve on a daily basis all studies published on COVID-19. From included studies we extracted: bibliometric parameters, country of studies' implementation and study design. We assigned papers to 25 a priori defined COVID-19-related topics and we described scientific outputs in relation to countries' academic publishing ranking, as well as COVID-19 burden. *Results:* 10,000 scientific articles were published on COVID-19 between 20th January and 7th May 2020, accounting for 2.3% of total scientific production over the study period. One third (33%) focused on COVID-19 clinical management, with little adherence to identified research priorities. Over sixty per cent of papers were opinion pieces not reporting original data. Papers were published on 1881 different journals but with half of scientific production included in 8% of journals. The US accounted for one fourth of total scientific production, followed by China (22.2%) and Italy (9%). *Conclusions:* Never before in the history of academic publishing such a great volume of research focused on a single topic, this being likely to introduce major changes in the way science is produced and communicated, at the risk of bringing it far from its ultimate aim: informing clinical and public health practice and decision making. (www.actabiomedica.it)

Key words: COVID-19, public health, research priorities, bibliometric analysis, academic publishing, scientometric analysis

Introduction

COVID-19 pandemic is sparking an unprecedented wave of publications. The Economist positively commented on such exponential publishing trends, underlining how the virus “has changed the way scientists do their work and talk to each other” (1). Experts claim COVID-19 has accelerated much anticipated cultural shift in biomedical publishing, the question

being, is it happening “for the best”? The global public health emergency context (2-4) has pushed, for example, journals in laudable efforts (5) to fast track peer reviews, and publishers to waive publication fees and provide free access to articles' content; it has pushed forward the preprints' model carrying both pro and con arguments. Do we, the scientific community, agree that “for the best” means to truly inform clinical and public health practice and decision making? In this

context we aimed to conduct a systematic screening, quantitative assessment and critical appraisal of the first 10,000 scientific papers published on COVID-19.

Methods

The “COVID-19 literature task force”, was established within the Clinical Epidemiology Unit of San Raffaele Research Institute in Milan (Italy) to inform and support decision making of hospital clinical teams and regional health authorities. A comprehensive research strategy was developed to systematically retrieve, on a daily basis, all scientific papers on COVID-19 indexed on Medline. All study designs were included, and no language restrictions were set. Selected variables of interest were extracted from included studies by two researchers in blind. An *ad hoc* algorithm was developed to extract from each paper all bibliometric parameters. We assigned each paper to the country of the first author’s affiliation. Retrieved records were manually classified by study type and assigned to one of 25 different *a priori* defined COVID-19 sub-topics. For clinical papers the specific clinical field of investigation was identified. Scientific journals’ impact factor was obtained from the Web of Science Journal database. We conducted descriptive analysis on all parameters and variables of interest; scientific production by topic, and sub-topic was assessed over time. Research production by country was compared to the total research volume for the year 2019 reported by the Nature Index (6). Scientific outputs by country were assessed against countries’ COVID-19 epidemiological burden.

Results

Our search strategy retrieved 12,201 PubMed indexed papers between December 31st 2019, and May 7th 2020. After removing duplicates and articles not related to COVID-19, 10,000 publications were selected for analysis. The first papers were indexed on PubMed on 20th January; Over the study period COVID-19-related papers accounted for 2.3% of all global scientific production, this percentage increasing over time, from

0.3% in February, to 7.1% in May. On average, 92.6 new papers on COVID-19 were indexed every day, with a peaks at 500 articles indexed per day in May. Included papers were published on 1881 different scientific journals. Less than 8% of journals published half of total COVID-19 scientific production, with two journals publishing over 200 papers on COVID in a bit more than three months (BMJ n=337, the Journal of Medical Virology n=230). On the contrary, 43.3% of journals only indexed one COVID-19 article each. Globally, scientific journals had a mean impact factor of 8.4. Half of indexed papers were published in journals with an impact factor lower than 3.5.

The largest share of papers published on COVID-19 focused on clinical aspects; in particular, 10.1% of total production was on clinical management of COVID-19 patients and 22.9% on other diseases and fields of medicine in relation to COVID-19 implications (Figure 1), with particular reference to oncology (11.4% of all specialty-specific papers) and cardiology (11.3%). 9.7% of papers reported on COVID-19 epidemiology or surveillance data in different national, regional and local settings, other explored topics included health services delivery (6.3%) and disease natural history (6.4%). Little has so far been published on new therapies and treatment evaluation (4.4%), although with increasing trends, on the contrary, data on the management and impact of containment measures decreased over the study period (Figure 2).

Table 1 reports indexed COVID-19 related papers, by country, together with countries’ COVID-19 burden (COVID-19 deaths from official surveillance data) and overall scientific production ranking. The largest amount of COVID-19 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, followed by China (22.2%). Italy was the country in Europe with the highest number of published papers on COVID (9%), followed by the UK (7.6%) and France (3.2%).

Discussion

We reflect on content, trends and quality of scientific publishing on COVID-19 since the start of the

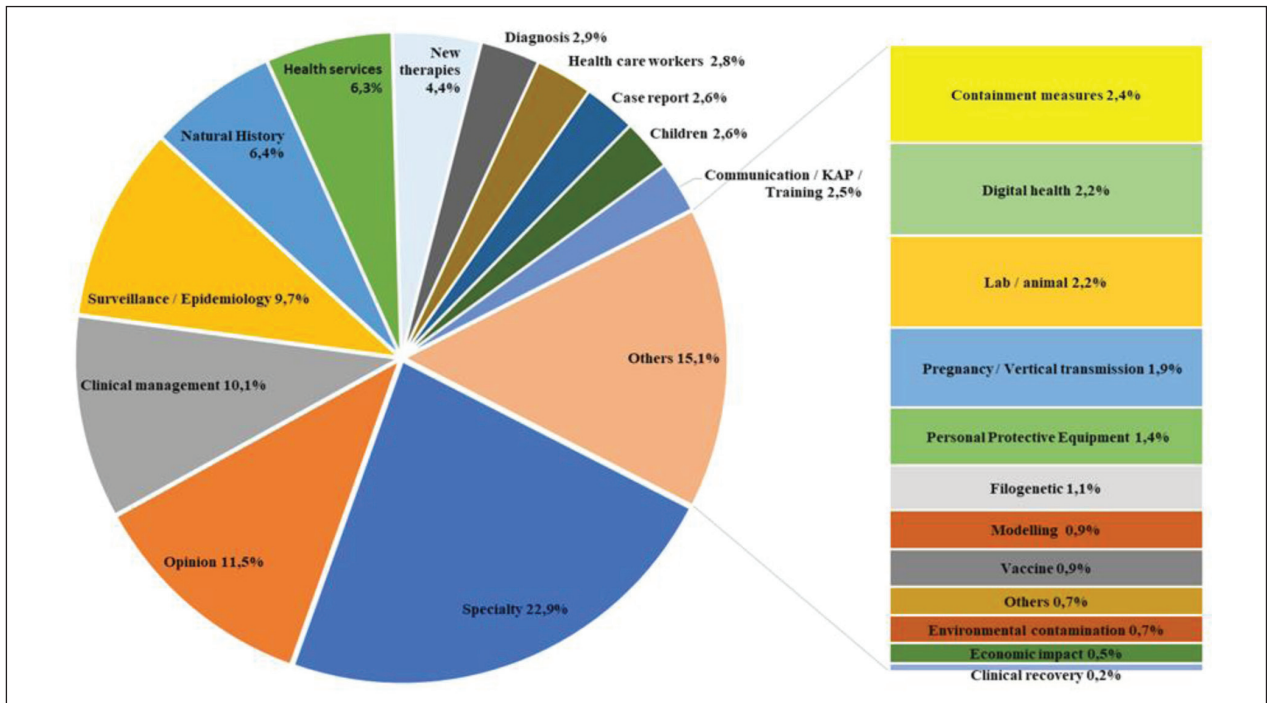


Figure 1. Distribution of total COVID-19 scientific output, by topic



Figure 2. Percentage of scientific papers published on COVID-19 containment measures and new therapies, over total COVID-19 scientific production, over time

Table 1. Country-level Covid-19 scientific production (n. of Pubmed indexed papers, 20th January-7th May), scientific production ranking (Nature Index for the year 2019) and COVID-19 burden (n. of COVID-19 deaths on 7th May from official statistics)

Rank	Country	COVID-19 Pubmed indexed articles (n.)	COVID-19 deaths	Nature index articles count (4)	Nature Index Rank (4)
1	USA	2,647	65,197	28,403	1
2	China	2,217	4,643	18,026	2
3	Italy	903	29,684	2,638	11
4	UK	764	30,076	7,837	4
5	France	315	25,769	5,054	5
6	India	284	1,783	1,655	15
7	Canada	223	4,111	3,408	7
8	Iran	176	6,418	293	40
9	Germany	161	7,119	8,770	3
10	Singapore	157	20	1,254	19
11	Australia	145	97	2,986	9
12	Spain	138	25,857	2,767	10
13	Brazil	130	7,921	856	24
14	Korea	123	256	2,419	12
15	Switzerland	103	1,504	3,270	8
16	Taiwan	88	7	931	23
17	Japan	83	551	4,905	6
18	Netherlands	73	5,204	2,403	13
19	Turkey	68	3,584	374	37
20	Thailand	67	55	254	42
21	Belgium	54	8,339	1,176	20
22	Saudi Arabia	51	209	455	30
23	Israel	45	238	1,291	18
24	Egypt	38	69	195	46
25	Ireland	37	1,375	450	31
	Tot 25 countries	9,090	230,086	102,070	-
	Rest of the World	571	23,959	-	-
	Missing country	339	-	-	-
	World Total	10,000	254,045	-	-

pandemic. Overall, we report that a large share of total scientific production in the first months of 2020 concentrated on COVID-19, with exponentially increasing trends. COVID-19 research so far focused on selected, mostly clinical topics and more than half of scientific publications do not include analysis of original data but are opinion pieces.

Research priorities on COVID-19 have been identified over time: at the end of January 2020, two days after the first-ever paper on COVID-19 (called 2019-nCoV at the time) was indexed in Pubmed, *Nature* listed the six questions scientists were (should have been) asking on COVID-19 natural history, epidemiology, sequencing, treatment and containment measures (7). Later in March, a much detailed list of epidemiological research priorities on transmission dynamics, the severity of the disease, immunity, and impact of control and mitigation measures, among others was identified as essential to decision making (8). However, our analysis, in line with our assessments on the literature on COVID-19 (9), demonstrate that little of the research published on COVID so far matches those identified priorities, leaving many key research questions still unanswered. Although we acknowledge longer times are needed for research to be conclusive on many aspects, we notice with concern that opinion papers, speculations and personal perspectives largely surpass reporting of original data. Even prior to the retraction of two papers from *The Lancet* and the *New England Journal of Medicine* (10, 11), some scientists had warned against the risk of dissemination of inaccurate and exaggerated information (12). We fear that not only the current pandemic has introduced major changes in the approach to academic publishing but also that its drivers might be moving away from science's ultimate aims to support the pursuing of individuals' and population wellbeing. Although our study has some limitations that need to be acknowledged, including having considered only studies indexed on Pubmed and not having extracted from all included studies details on methods and finding, which was outside the scope of our analysis, we provide a clear and comprehensive overview of what has been published on COVID-19 suggesting that the strong push for scientific dissemination and visibility around COVID-19 might not be going hand in hand with the production

of solid evidence much needed to inform and support clinical and public health decision making in these difficult times (13).

Conclusions

The scientific production on COVID-19 so far has the characteristics of "an epidemic in an epidemic". COVID-19 scientific output has gone viral, with both positive and negative implications. We would expect to see the current situation to bolster the value and potential of scientific research in pursuing individuals' and population wellbeing, asking the scientific community to resist the temptation of papers' inflation at the expense of scientific reputation and credibility.

Conflict of interest: Each author declares that he or she has no commercial associations (e.g. consultancies, stock ownership, equity interest, patent/licensing arrangement etc.) that might pose a conflict of interest in connection with the submitted article

References

1. The Economist. High-speed science. Available: <https://www.economist.com/leaders/2020/05/09/speeding-up-science-during-the-pandemic> [Accessed: 17.05.2020].
2. Odone A, Delmonte D, Scognamiglio T, Signorelli C. COVID-19 deaths in Lombardy, Italy: data in context. *The Lancet Public health*. 2020;5(6):e310.
3. Signorelli C, Scognamiglio T, Odone A. COVID-19 in Italy: impact of containment measures and prevalence estimates of infection in the general population. *Acta bio-medica : Atenei Parmensis*. 2020;91(3-s):175-9.
4. Amerio A, Bianchi D, Santi F, Costantini L, Odone A, Signorelli C, et al. Covid-19 pandemic impact on mental health: a web-based cross-sectional survey on a sample of Italian general practitioners. *Acta bio-medica : Atenei Parmensis*. 2020;91(2):83-8.
5. Vanelli M, Signorelli C, De Sanctis V. Foreword: Research in times of pandemic COVID-19. *Acta bio-medica : Atenei Parmensis*. 2020;91(2):11-2.
6. Nature. Nature Index. Available at: <https://www.natureindex.com/annual-tables/2020/country/all> [Accessed: 17.05.2020].
7. Callaway E, Cyranoski D. China coronavirus: Six questions scientists are asking. *Nature*. 2020;577(7792):605-7.
8. Cowling BJ, Leung GM. Epidemiological research priorities for public health control of the ongoing global novel coronavirus (2019-nCoV) outbreak. *Euro surveillance : bulletin European sur les maladies transmissibles = European communicable disease bulletin*. 2020;25(6).

9. 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]
10. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: Cardiovascular Disease, Drug Therapy, and Mortality in Covid-19. *N Engl J Med*. DOI: 10.1056/NEJMoa2007621. *The New England journal of medicine*. 2020.
11. Mehra MR, Ruschitzka F, Patel AN. Retraction-Hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet (London, England)*. 2020;395(10240):1820.
12. Ioannidis JPA. Coronavirus disease 2019: The harms of exaggerated information and non-evidence-based measures. *European journal of clinical investigation*. 2020;50(4):e13222.
13. Signorelli C, Odone A, Gianfredi V, Bossi E, Bucci D, Oradini-Alacreu A, et al. The spread of COVID-19 in six western metropolitan regions: a false myth on the excess of mortality in Lombardy and the defense of the city of Milan. *Acta bio-medica : Atenei Parmensis*. 2020;91(2):23-30.

Received: 25 June 2020

Accepted: 5 July 2020

Correspondence:

Prof. Anna Odone

Università Vita-Salute San Raffaele

Via Olgettina 60 - Milano

E-mail: Odone.anna@hsr.it