$9^{th}$ International Conference on Computational Social Science $IC^2S^2$
July 17-20, 2023, Copenhagen, Denmark

# The impact of COVID-19 on scientific productivity and collaboration

*Keywords: Scientific productivity, scientific collaboration, gender inequality, COVID-19, science of science*

## Extended Abstract

In early 2020, a novel coronavirus spread around the world, inducing lockdowns and severe restrictions on social interactions. Within a few days, it became clear that also science was experiencing one of the biggest disruptions since the second world war. Firstly, it had to quickly provide answers and solutions to the many issues raised by the new coronavirus emergency. Second, it had to do so under completely new conditions: scientists working full-time from their homes without face-to-face interactions, limited access to labs or working spaces, and no possibility to travel. The immediate and worrying effects of this unexpected disruption were brought to the attention of the scientific community [1]: young parents, caregivers, and especially female scientists experienced difficulties to continue doing research when schools closed [2, 3], students and early career scientists lamented isolation and uncertain career perspectives [1], and scientists in the "bench sciences" had to halt their research for long periods [3].

As the pandemic progressed, the way of conducting research evolved, with innovative methods of collaboration taking over that blend onsite and online activities. However, despite the adoption of this "new normal" for science, it is unclear how the pandemic has affected academic output in the long run. We address this question by looking at how the pandemic affected the overall scientific productivity of different groups in science, whether existing effects of inequality amplification from before 2020[4, 5] changed, and how it changed the way we collaborate. We answer these questions by analysing scientific papers indexed in the Microsoft Academic Graph (MAG) database. We focus on preprints since they have been a prominent and fast way to communicate science during the pandemic [6] – they are made public as soon as the research is complete –, and have been widely used to measure the change in publishing output during the pandemic [6, 7].

We find that, in MAG, the number of preprints has been steadily increasing since 2015, from 6,000 monthly preprints in 2015 to 12,000 at the end of 2019 (Fig. 1(a)). In March and April 2020, the number of preprints jumped to 14,000 monthly and has been stagnating since then. However, it has grown faster than the total number of publications, going from 2% of documents in the MAG in 2015 to 10% in 2021 (Fig. 1(a) inset). The sudden increase in the number of preprints in March-April 2020 and the following stagnation seem to be mostly driven by preprints about COVID-19 (Fig. 1(b) inset): This number has jumped to 4000 in April 2020, and then declined steadily, confirming that preprints have been an important outlet to communicate results fast about COVID-19 during the initial part of the pandemic [8]. The number of authors publishing a preprint for the first time also has been steadily increasing since 2015, seeing a sudden rise at the start of the pandemic around March 2020 (Fig. 1(c)). Although the rate of these new authors continued on an increasing trend even after the pandemic, the proportion of female authors, which hovered around 30% in 2019, fall below 25% in 2021, and has since then fluctuated between these two values (Fig. 1(c) inset).

Gender inequalities also appear in the total number of authorships (defined as an instance of an author on a paper), and not only among new authors: a gender gap is ever-present since 2015 (Fig. 1(d)). This gap has been widening over time in absolute numbers and seems the largest after the onset of the pandemic. However, the proportion between genders shows that the gap is shrinking, as the fraction of female authors is growing from 15% in 2015 to around 20% in 2021 (Fig. 1(d) inset). When looking at authorships divided by career seniority (Fig. 1(e)), we find that most authorships pertain to authors that have a career of at least 10 years, while new authors with at most 1-year seniority have been increasing the fastest. We also track the individual productivity of active scientists – they have published at least one paper in any 5-year period in their career –, defined as the number of papers published in a given period. The average productivity of individuals has been steadily increasing since 2015, followed by a brief, sharp increase after the onset of the pandemic (Fig. 1(f)), and then falling to the level of 2017 at the end of 2020. However, female productivity decreases at a slightly faster rate than male (Fig. 1(f) inset).

The MAG preprint data set shows that, before the pandemic, the fraction of collaborators with whom an author has not collaborated before (i.e. new collaborators) was steady at around 4.5%. This value spiked around half a year after the onset of the pandemic to over 5% Fig. 2. Interestingly, since then, the fraction has kept decreasing and has reached much lower than pre-pandemic levels in 2022 at around 3.5%. Although female authors have been establishing more new collaborations than male peers, their decline is faster than men, and the gap is shrinking. Taken together, most effects shown here were already in place and can't be directly attributed to the pandemic, yet, they appear more complex than anticipated. For example, average individual productivity has increased and sharply decreased after the pandemic, affecting, on average, *all* scientists. In contrast, the amplification of gender inequalities in productivity has turned out to be relatively small. Though, publications or preprints capture only one component of scientific activity; more longitudinal and continuously updated data about other activities, such as outreach, conference participation, and peer review, is needed.

# References

1. Woolston, C. *et al.* Scientists count the career costs of COVID. *Nature* **599,** 331–334 (2021).
2. Minello, A., Martucci, S. & Manzo, L. K. The pandemic and the academic mothers: present hardships and future perspectives. *European Societies* **23,** S82–S94 (2021).
3. Myers, K. R. *et al.* Unequal effects of the COVID-19 pandemic on scientists. *Nature human behaviour* **4,** 880–883 (2020).
4. Huang, J., Gates, A. J., Sinatra, R. & Barabási, A.-L. Historical comparison of gender inequality in scientific careers across countries and disciplines. *Proceedings of the National Academy of Sciences* **117,** 4609–4616 (2020).
5. Larivière, V., Ni, C., Gingras, Y., Cronin, B. & Sugimoto, C. R. Bibliometrics: Global gender disparities in science. *Nature* **504,** 211–213 (2013).
6. Fraser, N. *et al.* The evolving role of preprints in the dissemination of COVID-19 research and their impact on the science communication landscape. *PLoS biology* **19,** e3000959 (2021).
7. Abramo, G., D'Angelo, C. A. & Mele, I. Impact of Covid-19 on research output by gender across countries. *Scientometrics,* 1–16 (2022).
8. Waltman, L. *et al. Scholarly Communication in Times of Crisis: The Response of the Scholarly Communication System to the COVID-19 Pandemic* (Research on Research Institute, 2021), 10381325 Bytes.

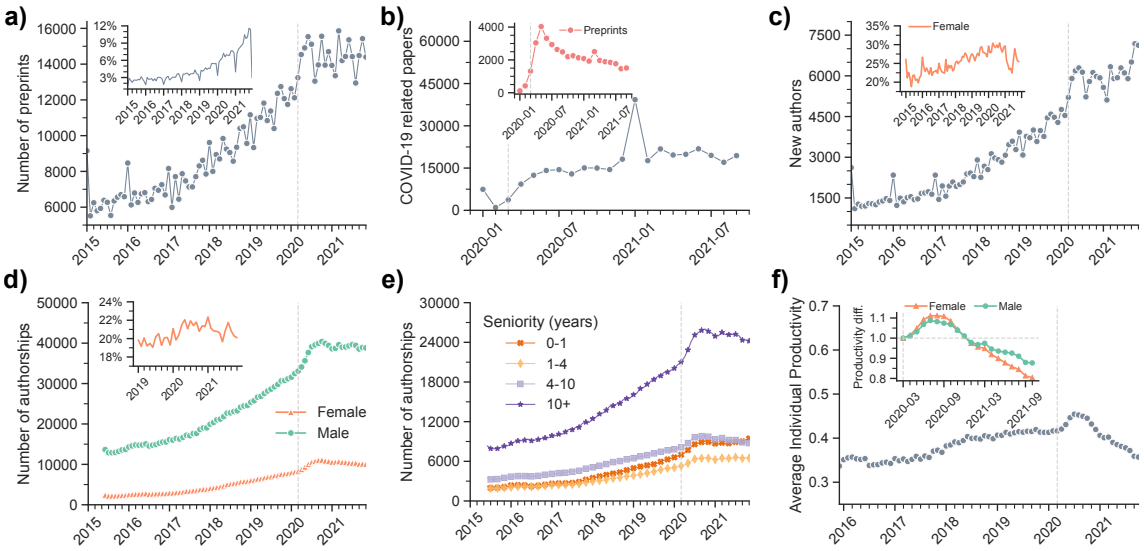


Figure 1: **Scientific productivity before and after the COVID-19 pandemic onset**. **a)** Monthly total number of preprints indexed by MAG. **Inset:** preprint proportion among all documents in MAG. **b)** Monthly number of COVID-19 related publications in OpenAlex after December 2019. **Inset:** Number of COVID-19 related preprints among all documents in OpenAlex. **c)** Monthly number of new authors, that is authors publishing a preprint for the first time. **Inset:** Proportion of new female authors. Note that the proportion of male authors is equal to 100%- percentage female new authors, since we excluded the authors with unidentified gender. **d)** Number of authorships stratified by gender. We display a rolling average with a 6 months window. **inset**: Proportion of female authorships. The proportion of male authorships is equal to 100%- percentage female new authorships. **e)** As in **d)**, with authorships stratified by seniority. Seniority is defined as the number of years between an author's current publication and their first one ever. **f)** Average individual productivity of active scientists, defined as the number of papers published by an individual scientist in a given time window. A scientist is considered active if they have published at least one paper in any 5-year period in their career. In the plot, we use a 6 months window. **Inset:** Difference between male and female individual productivity relative to March 2020, defined as the ratio of the gendered individual productivity difference and its value in March 2020.

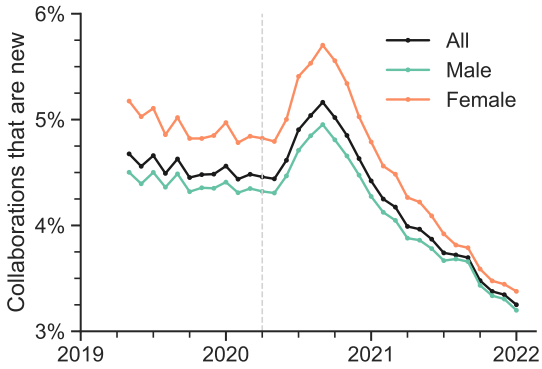


Figure 2: **Scientific collaborations before and after the COVID-19 pandemic onset**. At every point in time, we report the fraction of new collaborations in the previous 2-years, i.e the number of collaborators with whom an author has not collaborated before this 2-years window, divided by the total number of collaborators in this window. We report this measurement for all authors for which we could infer a gender (black line), and separately for female and male authors.