





Impact of the COVID-19 pandemic on publishing in astronomy in the initial two years

Received: 29 March 2022

Accepted: 13 October 2022

Published online: 28 November 2022

 Check for updates

Vanessa Böhm ^{1,2,4}  & Jia Liu ^{3,4} 

The COVID-19 pandemic and associated lockdowns changed working conditions for many researchers worldwide. While there exists initial evidence that these conditions have had a measurable impact on the field of astronomy, a comprehensive quantitative analysis is still outstanding. We study the effects of the pandemic on the astronomy community worldwide, with a special focus on early-career and underrepresented female scientists, using public records of publications. We find that the overall output of the field, measured by the yearly paper count, has increased. This is mainly driven by boosted individual productivity in most countries. However, a decreasing number of incoming new researchers is seen in most countries we studied, indicating higher barriers for new researchers to enter the field or complete their first project during COVID. The overall improvement in productivity is not equally shared by women. A smaller fraction of papers are written by female astronomers and fewer women are among incoming new researchers as compared to pre-pandemic trends, in 14 out of 25 countries we studied. Even though female astronomers became more productive during COVID, the level of improvement is smaller than for men. Pre-COVID, female astronomers in countries such as the Netherlands, Australia and Switzerland were equally as or even more productive than their male colleagues. During COVID, on average, no single country's female astronomers were able to be equally productive as their male colleagues.

The COVID-19 pandemic (hereafter COVID-19 or COVID for brevity) has impacted the personal and professional lives of people worldwide. Here, we study the impact of COVID-19 on the field of astronomy, using public records of astronomical publications from 1950 to February 2022, which includes 2 years during which research worldwide was disrupted by the COVID-19 pandemic. Astronomy has well-organized and digitalized publication records, making it possible to trace historical trends. Most articles are published as preprints with little delay between the time of production and the time of appearance on the

database. In addition, astronomy follows a by-contribution author ordering, making it relatively straightforward to quantify individual authors' contributions.

In addition, we study the impact of COVID-19 on gender gaps in astronomy. Pre-COVID, large gender gaps already existed in the science, technology, engineering and mathematics fields¹ and astronomy was no exception^{2–5}. The long-term causes for women to quit science, technology, engineering and mathematics jobs—workplace discrimination, lack of encouragement to advance their careers, lack of support in

¹Berkeley Center for Cosmological Physics, University of California, Berkeley, CA, USA. ²Lawrence Berkeley National Lab, Berkeley, CA, USA. ³Kavli IPMU (WPI), UTIAS, The University of Tokyo, Kashiwa, Japan. ⁴These authors contributed equally: Vanessa Böhm, Jia Liu. [✉]e-mail: jia.liu@ipmu.jp; vboehm@berkeley.edu

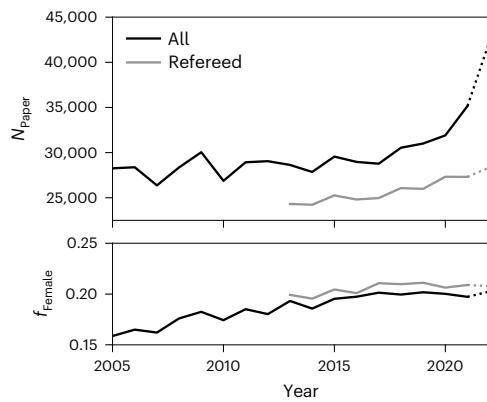


Fig. 1 | Number of papers in astronomy per year. In the top panel we show all publications in black and refereed publications in grey. We only include refereed papers for the past 10 years, as the journal review process usually takes only a few months. In the bottom panel we show the fraction of papers written each year by female first authors. The dotted lines are projections using partial-year data from 2022. The Poisson errors are at the sub-per cent level (nearly invisible if shown), and hence the fluctuations seen here are likely non-random.

balancing work and family, and lack of role models⁶—might have been amplified during COVID (see, for example, ^{7–11}). Earlier work by ref. ¹² on US- and Europe-based scientists showed that female scientists and scientists with young children are disproportionately affected by COVID-19, resulting in 5% and 17% larger declines than men in research time, respectively. Despite the fact that institutes have relaxed their rules to help early-career scientists cope with COVID-19, ref. ¹³ reported that many intended gender-neutral responses, such as work-from-home provisions and extensions on evaluations, may in fact have exacerbated underlying gender inequalities in academic sciences, engineering and medicine.

Finally, we study the impact of COVID-19 by country. Policy makers in different countries have implemented drastically different responses to the pandemic, such as stay-home orders, border restrictions, mask wearing, testing and contact tracing, vaccination, financial support and relief, and school and workplace closures¹⁴. We expect these measures to have varied impacts on the local scientific community. Within the Italian astronomical community, the submission to arXiv from women is significantly under average for 2020 with respect to the previous years, while that from men is larger by up to 10% (3.5σ)¹⁵. Similarly, women, researchers on short-term contracts such as PhD students and postdocs, parents and expatriates are particularly impacted by COVID-19 in the French astronomical community¹⁶.

Overall output in astronomy

We show in Fig. 1 the overall output in the field of astronomy, quantified as the total number of papers per year. The top panel shows the total number of all (refereed + non-refereed) papers per year since 2005. For 2022, we project the full-year output by dividing the number of papers published so far by the fraction of the year passed at the time of data download, shown as dotted lines. We note that publications in astronomy likely have seasonal effects due to job seasons, grant application deadlines and school calendars, so the simple projection here should be taken with caution. The Poisson errors are at the sub-per cent level (nearly invisible if shown), and hence the fluctuations seen here are likely not random.

We see a general trend of an increasing number of papers several years before COVID-19, roughly from 2014. The first 2 years with COVID—2020 and 2021—saw no decrease in the number of papers. We verify this trend by analysing only refereed papers, shown as the grey line. We only include refereed papers for the past 10 years, as the journal review process usually takes only a few months. For refereed

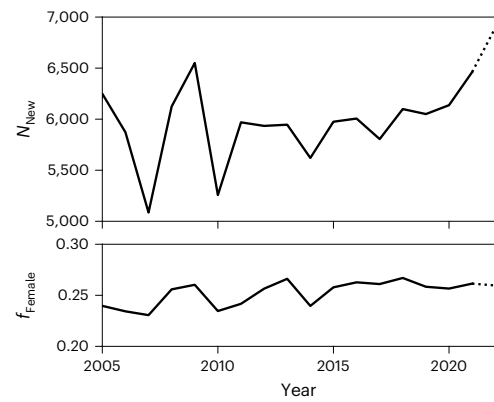


Fig. 2 | Number of new authors in astronomy per year. In the top panel we show the total number of new authors and in the bottom panel the fraction of women among them. The dotted lines are projections using partial-year data from 2022. The Poisson errors are at the per cent level (nearly invisible if shown), and hence the fluctuations seen here are likely not random.

papers, which are roughly 80% of all papers, the upward trend remains, though it is milder. Considering that many authors who submitted their papers to journals did not post their preprints, and the likely delayed referee process during COVID, the refereed curve likely reflects the lower limit. Therefore, we conclude that the field of astronomy has been more productive during COVID than pre-COVID.

In the bottom panel of Fig. 1, we show the fraction of papers written by female first authors. Because we have a large number (30%) of gender-identified authors, we compute the fraction of female authors using $f_{\text{Female}} = N_{\text{Female}} / (N_{\text{Female}} + N_{\text{Male}})$ instead of $N_{\text{Female}} / N_{\text{total}}$. The fraction of woman-led papers (~20%) has seen little to no change in the past 10 years, demonstrating a persistent gender gap in astronomy.

The increase in overall output in the field could be due to an increase in new researchers and/or an increase in individual productivity. We investigate the exact sources of the increase below.

Incoming new researchers

We expect that COVID-19 has had a more severe impact on junior researchers. Early-career researchers (students, postdocs and pre-tenure faculty) tend to have a smaller scientific network, fewer job securities and are more likely to have young children. They are hence more vulnerable than more senior, especially tenured, researchers during a crisis. We study the number of incoming new researchers each year. We define new researchers as those who publish their first paper in a given year, disregarding their author order. Our data do not distinguish the career stages at publication.

We show the number of new authors per year in the top panel of Fig. 2. The Poisson errors are at the per cent level, and hence any fluctuations seen are likely associated with socioeconomic changes. Compared to before 2020, we see an increase in new authors in recent years. However, as we will show later, the increase during COVID is largely driven by only a handful of countries in Asia. The majority of countries, including the USA and all European countries, experienced a drop in the number of new authors during COVID.

We show the fraction of women among all new authors in the bottom panel of Fig. 2. We again see no improvement for the past 10 years and possibly longer. This indicates strong barriers for women to enter the field of astronomy, likely due to barriers faced earlier during their education.

Individual productivity

We next study the impact of COVID-19 on the productivity of active researchers on an individual basis. We define ‘active’ as anyone who published in a given year. We quantify the level of productivity by counting

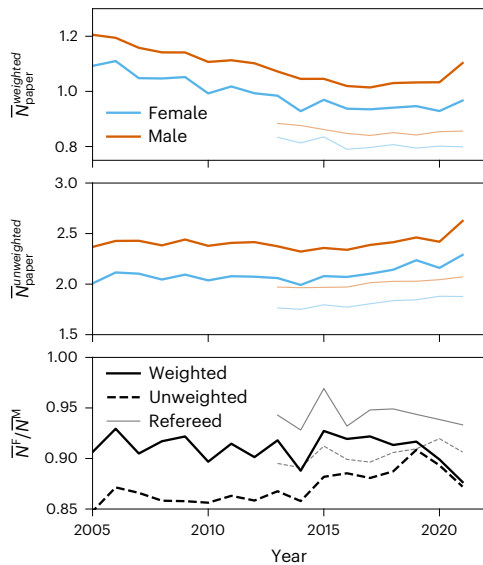


Fig. 3 | Average number of papers per author by year. We show the average number of papers active authors in our sample write each year, for both female (blue) and male (orange) authors. Active authors are defined as anyone who published in a given year. In the top panel the average is weighted by the position in the author list; in the middle panel each paper contributes equally to the average. In the bottom panel we show the ratios between the number of papers published by female and male authors (NF and NM, respectively). Lighter-coloured curves consider refereed papers only.

the number of papers each researcher authored, both weighted by their ranking and unweighted. Paper counting does not take into account the quality or impact level of the paper, as most papers published during COVID have yet to be widely cited. We estimate the workload each author has performed to contribute to a publication using

$$N_{\text{paper}}^{\text{weighted}} = \sum_i \left(\frac{1}{2} \right)^{(N_p^i - 1)}, \quad (1)$$

where N_p^i is the author’s position in the author order and i loops over all papers an author has written in a given year. In other words, we assume none-first authors spend roughly half of the effort of the previous author.

In addition, we measure unweighted publication counts $N_{\text{paper}}^{\text{unweighted}}$ for each active researcher. In this scheme, we count all papers equally, excluding papers with more than 16 authors. Combined with the weighted counts from above, unweighted counts provide additional information about collaborative level. An $N_{\text{paper}}^{\text{unweighted}}$ that is larger than an $N_{\text{paper}}^{\text{weighted}}$ shows that an author played a supporting role in multiple collaborative works—likely the case of senior researchers working with many junior students and postdocs. We show in Fig. 3 both weighted (top) and unweighted (middle) counts for female (blue) and male (orange) authors. We also show refereed papers only in lighter colours. For all papers (solid lines), we see an overall increase in both weighted and unweighted counts during COVID, though an initial drop is visible at the onset of COVID-19 in 2020, except for male authors’ $N_{\text{paper}}^{\text{weighted}}$. When considering refereed papers only, the increase during COVID is much more minor and nearly invisible in the figure. We conclude that active (female or male) authors’ productivity and scientific collaboration have not been reduced by COVID-19. An improvement may even be present, though longer-term data will be necessary to confirm this.

In the bottom panel of Fig. 3, we show the gender gap in productivity. Female authors are producing ~10% fewer papers than male authors, at all times (black solid lines). This persisting trend hints at social and systemic discrimination impacting the scientific productivity of

Table 1 | Excess productivity during COVID: we measure the excess productivity, EP, in terms of first-author publications (columns 2–4) and all publications (columns 5–7) during COVID as compared to the average productivity in the years 2010–2019. We report results separated by gender and career stage. Standard errors are computed by jackknifing on the pre-COVID data

Career stage	First-author publications			All author publications		
	$\frac{EP}{\sigma_{EP}}$ Male	$\frac{EP}{\sigma_{EP}}$ Female	$\frac{EP_{\text{Male}}}{EP_{\text{Female}}}$	$\frac{EP}{\sigma_{EP}}$ Male	$\frac{EP}{\sigma_{EP}}$ Female	$\frac{EP_{\text{Male}}}{EP_{\text{Female}}}$
1–6 yrs	2.73	1.74	1.72±1.17	2.16	1.55	1.51±1.20
6–11 yrs	1.63	1.54	1.55±1.39	1.66	1.56	1.50±1.32
11–16 yrs	1.74	0.49	3.46±7.30	3.17	1.76	1.56±1.01
16–21 yrs	0.54	-0.37	-0.98±3.25	1.69	0.81	1.37±1.86

women¹. In our by country analysis, we will show that female researchers in some countries consistently published more than or at similar rates as male researchers pre-COVID, demonstrating that female researchers are indeed as productive as men when provided with adequate support. Furthermore, the gender disparity is worse when taking collaborative work into consideration (dashed lines). This trend could be due to the fact that female authors are at earlier career stages on average and still need to establish their scientific network, but it could also point towards barriers to networking and joining collaborative research. The trend remains in the refereed-papers-only analysis (lighter colours), though with a slightly milder gender disparity. Female astronomers’ productivity (weighted paper counts) dropped from 92% pre-COVID to 89% during COVID, compared to that of male astronomers. The drop for unweighted paper counts is smaller, from 89% pre-COVID to 88% during COVID. The drop is less obvious for refereed papers, possibly due to a delayed referee process during COVID.

Increase in productivity by career stage and gender

We further study the interesting phenomenon of increase in productivity during COVID, by separating authors by their career stage. We compare the average productivity of researchers during COVID (defined as the time interval from 1 February 2020 to 31 January 2022) to the average productivity during similar time windows before COVID.

We consider a total of 5×2 -year time intervals from 2010 to 2019, which we treat as control experiments. From the pre-pandemic time intervals we measure the average productivity of active male and female researchers at different career stages. We define ‘active’ as having published either during or in the last 2 years before the time interval under consideration. Here we summarize the result by means of excess productivity, EP, where EP is defined as the difference between productivity during COVID and average productivity in similar time intervals pre-COVID,

$$EP = \overline{N}_{\text{paper}}^{\text{COVID}} - \langle \overline{N}_{\text{paper}}^{\text{pre-COVID}} \rangle, \quad (2)$$

where $\langle \dots \rangle$ averages over the five pre-COVID time intervals. Standard errors are computed by jackknifing on the pre-COVID data. We report the EP divided by the estimated standard error σ_{EP} for different career stages and genders. We further compute the ratio of the excess productivity between male and female researchers. Results are listed in Table 1 for both first-author publications and all publications.

We find a general increase in productivity. However, this increase is not equally shared among different genders and career stages. Earlier career stages seem to have seen the greatest increase in productivity, but also exhibit greater (more significant) disparities between men and women. We use both refereed and non-refereed publications in this

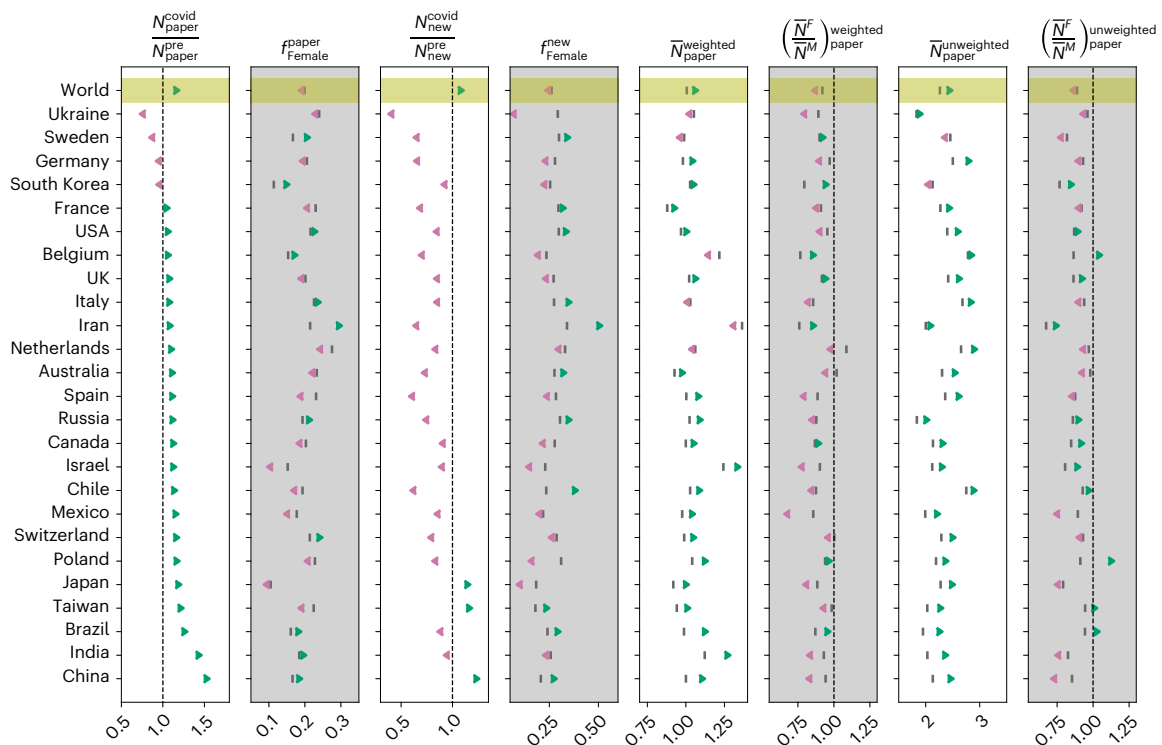


Fig. 4 | Impact of COVID-19 on astronomy by country. We measure the impact in terms of four statistics (white columns), each paired with a corresponding gender-disparity measurement (the adjacent grey column to the right). Grey points are pre-COVID and purple left-pointing (green right-pointing) triangles are during-COVID values that are worse (better) than pre-COVID. The world statistics are shown in the first row. Columns 1 and 2 (overall output): ratio of the number of papers per year during COVID to pre-COVID, and corresponding fraction of female authors. Columns 3 and 4 (new researchers): ratio of the number of new authors per year during COVID to pre-COVID, and corresponding fraction of new female authors. Columns 5 and 6 (individual productivity):

weighted average number of papers per active author $N_{paper}^{weighted}$, where none-first author papers are downweighted, and the corresponding ratio of female to male authors. Columns 7 and 8 (individual productivity irrespective of author position): unweighted average number of papers per active author $N_{paper}^{unweighted}$, where the author order is ignored, and the corresponding ratio of female to male authors. We only consider countries with more than 1,000 authors (25 in total). The error bars are not shown for legibility. For guidance, the Poisson errors are between the 1% level for the countries with the most authors and the 10% level for the ones with fewer authors.

analysis, since we expect a several-month delay between production and completion of the referee process. The observed trend is the same for refereed articles, but less significant.

By country

Finally, we study the same statistics for authors in different countries. In response to the pandemic, governments in different countries implemented drastically different policies. Analysis by country could reflect the impact of different COVID responses on the scientific community. We use the country of the most recent affiliation for each author. There are, in total, 141 countries in our database. We only include countries with more than 1,000 identified authors, to ensure that Poisson noise is at most at the 10% level, resulting in a total of 25 countries. The selected countries cover a wide range in geography, culture, economic development, social welfare, scientific priority and COVID-19 policies. We compute the same statistics as above (for the entire world) pre- and during COVID for each country. For pre-COVID statistics, we use the average of 5-year data immediately before COVID (2015–2019). For COVID statistics, we use the average of 2-year data during COVID (2020 and 2021). Figure 4 shows the four statistics—overall output, number of new authors, individual productivity weighted by author position and individual productivity irrespective of author position—in ratios of post- to pre-COVID. Each statistic is shown in pairs of white and grey columns, for the general population and the corresponding gender disparity measurement, respectively. We rank-order the countries by the first column (overall

output) and show the world’s statistics in the first row (shaded in yellow) to guide visual comparison. Grey points are for pre-COVID measurements, while purple (green) triangles show during-COVID values that are worse (better) than pre-COVID. We also show the fraction of female researchers who are active between 2015 and 2022 in each country in Supplementary Table 1.

In the first column, which shows the ratios of the total number of papers per year post- to pre-COVID, the majority of countries see an increase in scientific output during COVID (except for a few countries, notably Ukraine and Sweden). To identify the drivers of this increase, we show in column 3 that most countries see a decreasing number of new authors compared to pre-COVID, and hence incoming new authors are not responsible for the increased output worldwide. Rather, we see improved individual productivity for most countries (columns 5 and 7), driving up the worldwide paper output.

Despite the overall increase in scientific output, more than half of the countries see a worsened situation for women. Column 2 shows the fraction of papers written by female first authors pre- and during COVID. In general, only 20% of papers are written by female first authors. Column 2 can also be considered a rough indicator of the fraction of female researchers in each country. During COVID, 14 out of 25 countries saw an even smaller fraction. The exact reasons for this trend are not deducible from our analysis; we observe declines in the scientific output of female researchers in both better-than-average countries (for example, the Netherlands) and worse-than-average countries (for example, Japan and Israel).

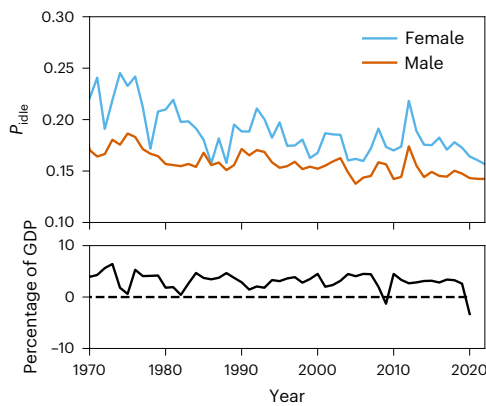


Fig. 5 | Fraction of idle authors per year. Idle authors are defined as authors who haven't published in the past 2 years but were previously active authors, defined as producing $N_{\text{paper}}^{\text{weighted}} > 1$ in the 2 years before becoming idle. Active female authors (blue) are more likely to become idle than active male authors (orange), at all times. In the bottom panel we show the world GDP growth¹⁷, to illustrate a possible anti-correlation of P_{idle} with socioeconomic changes. The COVID-19-induced recession is the only one (among six notable recessions) since 1970 without an associated peak in P_{idle} for either gender.

When examining the number of incoming new authors (column 3), we find that the increasing world average is mostly driven by only a handful of countries, namely China, Japan and Taiwan. Most other countries see a reduced number of new authors, dropping down to as low as 50–70% of the pre-COVID level. When examining the fractions of women among new authors in each country, we see that they remain roughly at pre-COVID levels (column 4), indicating no additional barriers for women to enter astronomy during COVID.

Finally, when studying individual productivity, we find that most countries see an improvement in individual productivity (column 5) and collaboration (column 7). However, this improvement is not equally shared by female researchers. Individual productivity has declined for women (when compared to men) in 17 out of the 25 countries. More strikingly, no single country's female researchers were able to be as productive as men during COVID, including the previously 110% outperforming female astronomers in the Netherlands. A similar situation is seen when taking into consideration collaborative work (column 8), where 13 out of 25 countries experienced a decline for women. These findings indicate that COVID-19 has taken a higher toll on female researchers.

Idling authors

Here, we turn our focus to the most vulnerable groups who are likely already adversely affected by COVID-19: the previously active authors who became idle during COVID. We show the probability for an active researcher to become idle in Fig. 5, for both female (blue) and male (orange) authors. We see that female authors have a much higher chance to turn idle than male authors, at all times.

We hypothesize that a 2-year idling period means that an author has likely left the field. To back this claim, we consider the normalized histogram of idle times, defined as the length of the period between two consecutive publications by the same author, in Fig. 6. Idle times are measured between the first and last publication that we have on record for each author. The normalized histogram of idle times represents the frequentist probability of observing an idle time. To put more emphasis on current publication trends and the younger author population, who are more likely to have young children, we made a cut to include only authors who had their first publication in the last 20 years. We verified that considering different or no cuts changes the results only by a few per cent and does not alter the conclusion. We find that the probability of an idle time of 2 years (that is, 2 years without writing a paper) is only

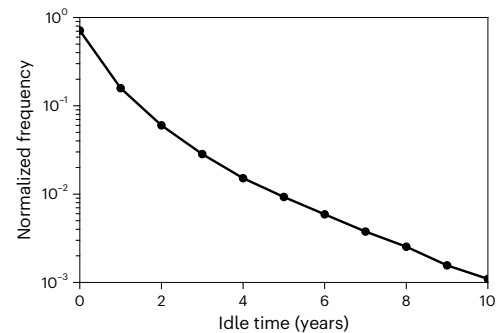


Fig. 6 | Normalized histogram of idle times. Idle time is defined as the time period between two consecutive publications and is measured in authors whose careers started less than 20 years ago (measured by year of first publication). We verified that considering different or no time cuts changes the results only by a few per cent points.

about 6% (6.4% for women and 5.9% for men). The probability of having an idle period of 3 years or longer (cumulative) is 7.0% (for both men and women). This suggests that an average author only has a ~13% chance of ever publishing again if they haven't published for 2 years. Therefore, an author who has not published in 2 years has likely left the field with a probability of 87% derived from historical data.

To investigate the potential correlation with the world economy, we show the world Gross Domestic Product (GDP) growth in the bottom panel¹⁷. We find almost one-to-one correspondences between past recessions (for example, 1975, 1980–1982, 1991–1993, 2001, 2008–2009) and peaks in P_{idle} for female authors. The trend is similar (though milder) for male authors. The recent (2008–2009) global financial crisis may have two associated peaks (2008 and 2012). We found significant anti-correlation between the world GDP growth and P_{idle} with Pearson's $r = -0.42$ ($P = 0.0023$) and $r = -0.47$ ($P = 0.0006$) for female and male authors, respectively.

Surprisingly, we do not see an increase in author idling during the current COVID-induced recession, in contrast to all previous recessions in the past 50 years. This could be due to the limited time period we consider, and a future peak might yet be to appear. However, previous P_{idle} peaks seem closely associated with corresponding crises (the peaks coincide with the GDP minima), and we even observe a mild drop in the idling rate during COVID. This seems to indicate that the impact of COVID-19 may be different to historical economic crises. Changes in the cultural and social aspects of astronomical research—for example, reduced commutes and business trips and increased flexibility in work modes—may have allowed otherwise busy researchers to focus on research and become more productive during COVID. This may also be partly due to the temporary extension of contracts implemented by institutes during COVID.

Discussion

COVID-induced school and work closures, hiring freezes, loss of child-care, increased eldercare duties and related mental health issues, among others, have had undeniably negative impacts on researchers in all fields. However, changes such as increased flexibility in work arrangements, reduced commutes and business trips, as well as improved virtual technologies, among others, have been potentially favourable for conducting scientific research. In studying the impact of COVID on the astronomical community, we found:

- The overall output of the field, measured in terms of the number of publications, has increased during COVID, for the whole world (by 13%), as well as for most countries (21 out of 25).
- Most countries (22 out of 25) have seen a decreasing number of incoming new researchers, except Japan, Taiwan and China,

indicating larger barriers during COVID for new researchers to enter the field or for junior researchers to complete their first project.

- Most countries have seen boosted individual productivity, both in terms of increased contribution to scientific papers and contribution to more papers (19 and 23, respectively, out of 25 countries).
- While the world has seen an improvement in researchers' productivity, the gender disparity has widened: a smaller fraction of papers were written by women, and women made up a smaller fraction of incoming new researchers in more than half of the countries we studied. Even though women were also more productive during COVID, the level of improvement is smaller than for men. For example, pre-COVID, female astronomers in the Netherlands, Australia and Switzerland were equally as or more productive than their male colleagues. During COVID, no single country's female astronomers were able to be more productive than their male colleagues on average.

Our study has several caveats. First, the underlying causation of the trends seen in our study can be complex and should be carefully discussed and studied in the future. Second, our gender-based analysis relied on tools designed for binary genders and hence overlooked the non-binary people in the community¹⁸, who might have been affected more by the COVID-19 pandemic. We note that other types of inequalities, such as those related to race, may also have been amplified during COVID, though this is beyond the scope of this work. Finally, we only have data from a limited period of time after the onset of the pandemic, but its impact is expected to be long term. While we are able to study the quantitative outputs of the field during COVID, the quality of these papers is yet to be seen. It would be valuable to revisit this topic in a few years.

Methods

Data acquisition

The dataset used in this study was obtained through the SAO/NASA Astrophysics Data System API in a single download on 13 February 2022. The downloaded data includes all entries on the Astrophysics Data System server of type 'article' or 'eprint' that were published between 1 January 1950 and 13 February 2022 and included in the 'astronomy' database. While most of our analysis focuses on recent publication patterns, we note that a long historical record is necessary to identify authors' career stages and that this increases the chance of identifying authors' full given names, which is crucial for our gender-related analysis.

We include both refereed and non-refereed articles (preprints) for the main analysis, since our primary goal is to measure productivity instead of scientific impact, which would be too early to judge for recent publications. Including preprints is also necessary to reduce potential biases due to the time lag between the completion of a paper and the acceptance to a journal, which we expect to have been prolonged during COVID-19. Where necessary, we verify observed trends by repeating our analysis on the subset of refereed articles.

We discard publications with more than 16 authors, since these are generally large collaboration papers with alphabetical author order. The cutoff at 16 authors is somewhat arbitrary; we would like to include as many papers as possible but exclude large collaboration publications, often with hundreds of authors. The cut at 16 authors corresponds to keeping about 95% of the dataset. We verified that the presented results are not sensitive to the exact location of this cut. After this cut, we obtain a total of 1,207,197 publications. For each publication entry, we obtain the title of the article, the publication date, the names of the authors and their affiliations.

Author identification

Next, we transform the publication dataset into an author dataset for all authors who have had at least one first-author publication. We used

first-authorship as a way to isolate individuals who have dedicated a considerable amount of their time to a scientific project in astronomy and can hence be considered part of the field. In this section, we mainly follow ref. 5's method to identify unique authors.

While the full family names are usually provided, the given names—including the first and middle names—are often in initials. We separate author entries into three categories: with full given names, with initial-only given names and without a given name. We first match entries with the same full given + family names. We then match initial-only names to that list by unique initial + family names combination. For initial-only names without a match in the full given name list, we create a new entry for each unique initial + family name; these entries are not included in the gender analysis, as we are unable to identify the gender based on initials. Finally, we discard any entry without any given name (neither initial nor full), as it is not possible to identify unique authors only by their family name with confidence.

Furthermore, we consider possible changes in family names, often associated with changes in family status such as marriage and divorce. To do so, for entries with compound family names—either joined by a hyphen or by a space—we search for single family name entries that match either of the individual family names. Once we find a match, we merge the entries if they also have the same given (full or initial-only) names. We also discovered that we are able to further match unique non-Latin names (such as Chinese, Japanese or Korean names), as some journals allow authors to publish their names in their own language alongside the English versions of their names. We confirm the validity of such a procedure by manually comparing other information from the matched entries, such as the paper subject and affiliation. One caveat is that a complete change in family name would not be captured in our method. This is unavoidable because it is impossible to identify authors merely by their given names. As a result, such authors will be split into two entries, and a more senior author would be mistaken as two junior authors. However, the entries with identified name changes correspond to only 0.4% of the full database, which suggests that the impact of undiscovered name changes is negligible.

Of a total of 639,068 author entries, 53% have full given names (337,449), 47% have initial-only given names (301,597) and < 1% have no given names (22). We are able to match 40% of the initial-only entries (120,371) to the full given name entries. Taking into account family name changes, we find an additional 2,681 matches. In total, we identify 516,304 unique authors, among which 258,889 have at least one first-author publication. We use the dataset with the 258,889 first authors for our analysis.

Gender identification

To study the impact of COVID-19 by gender, we identify the gender for each unique author entry (258,889 in total). We use the genderize.io API¹⁹, which assigns a gender and a probability (P_{gender}) to a given name. We note that this method does not permit us to identify non-binary genders—a limitation due to the data available to us. We acknowledge this shortcoming and that the binary classification assumed in this study does not fully capture the diversity of gender identities in the astronomy community.

When an entry has multiple given names, we examine the gender of each name. Usually they return the same gender, in which case we take the highest probability as P_{gender} . When the given names return different genders, we take the gender for the first given name if it has $P_{\text{gender}} > 80\%$ —this is to follow the tradition that the first given name, instead of the middle name, is more often used daily. If the first given name has a low gender identity ($P_{\text{gender}} < 70\%$), we take the name with the highest P_{gender} and assign that gender and P_{gender} to the entry.

We are able to identify the gender for 68% of entries (176,102). For the remaining entries with unidentified genders, most are initial-only names (30%) that are impossible to identify. The rest are rare names without a gender record (2%). We apply a cut of $P_{\text{gender}} > 80\%$ in our

gender-related analysis. This leaves us with 30,930 female entries (20%) and 126,529 male entries (80%), further reducing our completeness from 68% to 61%.

Anonymized author dataset

We take great care to anonymize the data as early as possible in the analysis process and remove all identifiable information. After author and gender identification, we remove the author names and replace them with unique author IDs. We do not keep any file that contains author names, but instead only record the gender that was assigned to each ID. We further replace the authors' affiliations with the countries of the affiliations.

After anonymization, we use the publication dataset to generate an author dataset. The author dataset contains the number of publications in each year between 1950 and 2022, as well as the author order and the country of affiliation for each publication. We conduct all our analyses with this condensed and anonymized dataset.

Monthly analysis

Excess productivity by gender and career stage during COVID (defined as the time interval between February 2020 and January 2022) is measured using the averages of five pre-pandemic time intervals as a counterfactual. We do not observe a general trend towards increasing or decreasing productivity pre-COVID that would require a more elaborate model.

To account for a possible dependence of productivity on career stage, we separate the data: for each time interval, we only consider publications from researchers who had their first publication within a certain period before the time interval under consideration. This ensures that we compare researchers at roughly the same career stage.

To correct for fluctuations in the number of active researchers in the field, we normalize the results by dividing the number of publications by the number of researchers who entered the field during a chosen time window, and who we also assume are still active during the time interval under consideration. We count authors as active if they published either during the 2 years before or during the time interval under consideration. We show results for four career stages, spanning a total of 20 years. We separate the results by gender. Figures showing the average monthly paper output per active author and career stage for all five 2-year time intervals are provided in the Supplementary Information as Supplementary Figs. 1 and 2.

Identifying idling active authors

To study previously active authors who became idle during COVID, we define this population to be those who did not publish any paper during the 2 years of COVID, but were productive ($N_{\text{paper}}^{\text{weighted}} > 1$) during the 2 years immediately before the pandemic (2018 and 2019). For comparison, we apply the analysis to as early as 1970, using data from 4 years before a specific year, in order to find historical trends potentially related to previous global crises. For a given year y , we define the active authors to be those who wrote $N_{\text{paper}}^{\text{weighted}} > 1$ during $[y - 3, y - 2]$. We then count the number of idling authors among them as those with no publications during $[y - 1, y]$. The fraction of idling active authors is the ratio between the latter to the former.

Data availability

The anonymized author dataset is available for download on Zenodo: <https://doi.org/10.5281/zenodo.7145196> (ref. 20).

Code availability

Code for reproducing results is hosted on a private Github repository and will be made available upon request.

References

1. UNICEF. Mapping gender equality in STEM from school to work. (UNICEF Office of Global Insight & Policy, 2020).

2. Davenport, J. R. A. et al. Studying gender in conference talks – data from the 223rd meeting of the American Astronomical Society. Preprint at arXiv 1403.3091 (2014).
3. Reid, I. N. Gender-correlated systematics in HST proposal selection. *PASP* **126**, 923 (2014).
4. Patat, F. Gender systematics in telescope time allocation at ESO. *Messenger* **165**, 2–9 (2016).
5. Caplar, N., Tacchella, S. & Birrer, S. Quantitative evaluation of gender bias in astronomical publications from citation counts. *Nat. Astron.* **1**, 0141 (2017).
6. Funk, C. & Parker, K. Women and men in STEM often at odds over workplace equity. *Pew Research Center* (2020).
7. Andersen, J. P., Wullum Nielsen, M., Simone, N. L., Lewiss, R. E. & Jagsi, R. Meta-research: COVID-19 medical papers have fewer women first authors than expected. Preprint at 2005.06303 (2020).
8. King, M. M. & Frederickson, M. E. The pandemic penalty: the gendered effects of COVID-19 on scientific productivity. *Socius* **7**, 23780231211006977 (2021).
9. Viglione, G. Are women publishing less during the pandemic? Here's what the data say. *Nature* **581**, 365–366 (2020).
10. Squazzoni, F. et al. Gender gap in journal submissions and peer review during the first wave of the COVID-19 pandemic. A study on 2329 Elsevier journals. *PLoS ONE* <https://doi.org/10.1371/journal.pone.0257919> (2021).
11. Abramo, G., D'Angelo, C. A. & Mele, I. Impact of COVID-19 on research output by gender across countries. *Scientometrics* <https://doi.org/10.1007/s11192-021-04245-x> (2022).
12. Myers, K. R. et al. Unequal effects of the COVID-19 pandemic on scientists. *Nat. Hum. Behav.* **4**, 880–883 (2020).
13. National Academies of Sciences, Engineering, and Medicine. *The Impact of COVID-19 on the Careers of Women in Academic Sciences, Engineering, and Medicine* (The National Academies Press, 2021).
14. Hale, T. et al. A global panel database of pandemic policies (Oxford COVID-19 government response tracker). *Nat. Hum. Behav.* **5**, 529–538 (2021).
15. Inno, L., Rotundi, A. & Piccialli, A. COVID-19 lockdown effects on gender inequality. *Nat. Astron.* **4**, 1114 (2020).
16. Leboulleux, L., Cantalloube, F., Choquet, E., Huby, E. & Singh, G. Impact of the COVID-19 crisis on the French astronomy community. In *SF2A-2021: Proceedings of the Annual Meeting of the French Society of Astronomy and Astrophysics* (eds Siebert, A. et al.) 252–265 (2021).
17. The World Bank. GDP growth (annual %) (2021). *World Development Indicators* <https://data.worldbank.org/indicator/NY.GDP.MKTP.KD.ZG> (2021).
18. Rasmussen, K. et al. The nonbinary fraction: looking towards the future of gender equity in astronomy. *BAAS* **51**, 75 (2019).
19. Demografix ApS. genderize.io <https://genderize.io/> (2022).
20. Boehm, V. & Liu, J. Anonymized author dataset from the publication “Impact of the COVID-19 pandemic on publishing in astronomy in the initial two years”. Zenodo <https://doi.org/10.5281/zenodo.7145196> (2022).

Acknowledgements

We thank M. Charisi, K. Choi, Z. Haiman, T. Hartwig, K. Inayoshi, Y. Ito, I. Okounkova, U. Seljak, D. Spergel, L. Yang and E. Schaan for their helpful discussions and comments. We also thank our anonymous referees for their constructive comments.

Author contributions

V.B. and J.L. planned the study, conducted the data analysis and wrote the paper with equal contributions.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information The online version contains supplementary material available at <https://doi.org/10.1038/s41550-022-01830-9>.

Correspondence and requests for materials should be addressed to Jia Liu or Vanessa Böhm.

Peer review information *Nature Astronomy* thanks Laura Inno, Ágnes Horvát and Lucie Leboulleux for their contribution to the peer review of this work.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Springer Nature or its licensor (e.g. a society or other partner) holds exclusive rights to this article under a publishing agreement with the author(s) or other rightsholder(s); author self-archiving of the accepted manuscript version of this article is solely governed by the terms of such publishing agreement and applicable law.

© The Author(s), under exclusive licence to Springer Nature Limited 2022