Surprising Resilience of Scientific Publication during a Global Pandemic: A Large-Scale Bibliometric Analysis

Rusti, Casandra Information Sciences Institute, University of Southern California, USA| mrusti@isi.edu
Ahrabian, Kian Information Sciences Institute, University of Southern California, USA | hrabian@isi.edu
Wang, Ziao Tsinghua University, Beijing, China | ang-za19@mails.tsinghua.edu.cn
Pujara, Jay Information Sciences Institute, University of Southern California, USA | jpujara@isi.edu
Lerman, Kristina Information Sciences Institute, University of Southern California, USA | lerman@isi.edu

ABSTRACT
Drawing on a global bibliographic corpus covering more than 23 million papers and 10 million disambiguated authors, we present the first longitudinal, institution-level portrait of how COVID-19 perturbed research activity and collaboration. Using multilevel regression and interrupted-time-series analysis, we trace participation, productivity, and collaboration for researchers at the 1,000 historically most-productive universities prior to 2020, stratified by geography, field, career stage, and gender. Publication counts and co-authorship networks surged in 2020, signaling an unexpected, rapid mobilization and resilience of the research system. Yet by late 2022 these metrics had reverted to their pre-pandemic trajectories, indicating that the spike was a short-lived reprioritization rather than a lasting shift. The lag inherent in many experimental pipelines – especially wet-lab science – raises the prospect of delayed losses not yet visible within our time frame. Our study establishes an evidence-based baseline for monitoring longer-term effects and offers actionable insights for science-policy makers seeking to safeguard research capacity during future global crises.

KEYWORDS
scientometrics; COVID-19; co-authorship networks; research productivity; gender

INTRODUCTION
The COVID-19 pandemic disrupted people's lives and work routines globally. Scientists and researchers at academic institutions were among those deeply affected by the pandemic due to the closure of colleges and universities, the move to online learning, and the shuttering of labs (Joseph, 2020). Studies published early in the pandemic warned of the adverse effect of the disruptions on research productivity, especially on female researchers and those with childcare responsibilities (Giurge, et al., 2021) (Myers, et al., 2020) (Morgan, et al., 2021) (Gao, et al., 2021).

Reductions in research productivity would slow scientific innovation precisely when new knowledge is most needed. In this study we treat research-system resilience as the capacity of the scientific community to absorb external shocks, sustain core functions, and return to its prior trajectory. We operationalize that capacity with three observable constructs: participation (retention and entry of active researchers), productivity (volume of published output), and collaboration (persistence of co-author networks that circulate skills and resources). A resilient system, therefore, is one that keeps scientists engaged, maintains or quickly restores publication levels, and preserves the collaborative ties that underpin knowledge creation.

Our study aims to analyze changes in researcher outcomes during the COVID-19 pandemic. Specifically, it characterizes three types of researcher outcomes representing metrics of scientific success: i) participation, ii) productivity, and iii) collaboration. *Participation* captures the number of scientists taking an active role in the scientific community, conducting research, and publishing papers to report research findings. *Productivity* encompasses the number of novel and impactful scientific discoveries resulting from research. *Collaboration* measures diverse researchers working together to create scientific advances. We posit that a resilient scientific community will continue to increase the participation of new researchers, maintain their productivity, and support collaborative teamwork that creates scientific innovation.

To measure the researcher outcomes of participation, productivity, and collaboration, we use data from a large-scale bibliographic catalog called OpenAlex (Priem, et al., 2022). OpenAlex contains information about hundreds of millions of publications and rich metadata about the time of publication, fields of study, authors and their affiliations, and references. We extract features from OpenAlex, such as unique authors, publication counts, and co-authorship edges, to develop several quantitative measures to capture research outcomes. By creating a time series for each of these measures, we can assess the impact of the COVID-19 pandemic on research success and determine the resilience of the scientific community.

Beyond characterizing the changes during the COVID-19 pandemic, our study explores the various factors that impact these research outcomes. We take advantage of the globally heterogeneous pandemic responses to disentangle pattern changes in research outcomes. During the pandemic, mitigation policies varied by country (e.g., China's "zero COVID" policy vs. Sweden's more hands-off approach), US state (e.g., strong vs. weak COVID-19 response), community (e.g., political climate and school closures), and even institution (e.g., remote work policy, vaccine mandates). The pandemic also had a different impact on researchers in different fields (biology vs. computer science), different career stages (junior vs. senior researchers), and different gender (Muric, et al., 2020) (Myers, et al., 2020). These diverse responses to the pandemic create conditions that allow us to identify patterns affecting the resilience of scientific innovation. To this end, we perform a temporal analysis of the correlation of institutional prestige, geographic region, seniority, gender, and fields of study on research outcomes.

Our study focuses on researchers affiliated with the 1,000 historically most-productive universities prior to 2020 and provides a stratified analysis based on the authors' institution. We hypothesize that the pandemic adversely affected researcher productivity and collaboration. To test this hypothesis, we operationalize measures of interest using bibliometric information and analyze these variables. Specifically, we pose the following research questions:

RQ1 How did researcher *participation* change during the pandemic at different research institutions world-wide?
RQ2 How did researcher *productivity* change during the pandemic?
RQ3 How did research *collaborations* change during the pandemic?
RQ4 Were there systematic differences in research outcomes based on the ranking and geographic region of institutions, researcher seniority and gender, and research publication's field of study?

Our examination of over 25 million publications from 10 million authors at these top 1,000 institutions has yielded an unexpected insight: the COVID-19 pandemic catalyzed a period of accelerated scientific activity. This finding stands in stark contrast to several smaller-scale studies conducted in the early stages of the pandemic that identified the detrimental impact of the pandemic on scientific productivity (Muric, et al., 2020) (Myers, et al., 2020) (Gao, et al., 2021). In this paper, we present a comprehensive analysis of the pandemic's effects on scientific research practices, enriching the understanding of resilience and adaptability within the academic sector amid a global crisis. To challenge prevailing assumptions and test assertions about the mechanisms by which the pandemic disruptions could affect scientific output, our study explores the factors shaping this period of intensified research activity. Our comprehensive analysis not only evaluates the immediate effects of the pandemic but also provides a foundation for understanding the long-term implications of such global disruptions on scientific progress.

RELATED WORKS

Scientific research requires efficient and robust infrastructure. Institutions, a category that includes universities, government labs, industrial labs, and national academies, provide this infrastructure (Taylor, et al., 2019) (Raan, 2013). Despite the long tradition of bibliometric and science of science research (Fortunato, et al., 2018), the focus has only recently shifted from individual scientists (Sinatra, et al., 2016) (Wang, et al., 2013) and teams (Wuchty, et al., 2007) (Milojevic, 2014) to analyzing how institutions affect researcher productivity, collaboration and impact (Way, et al., 2019) (Deville, et al., 2014) (Burghardt, et al., 2020). Still, many gaps remain in our understanding of the role of institutions in the resilience of scientific innovation and how they facilitate and support scientific collaborations through global and local disruptions.

Prior works tell us that the COVID-19 pandemic has taken a toll on parents, especially mothers (Giurge, et al., 2021). The increased childcare responsibilities due to pandemic-related school closures have led to a drop in the work hours of researchers, especially among researchers with young children (Myers, et al., 2020) (Morgan, et al., 2021) (Staniscuaski, et al., 2021). The work-life balance challenges have affected the productivity of women researchers, with fewer women becoming involved in COVID-19 research (Vincent-Lamarre, et al., 2020).

One study examined gender disparities in the authorship of tens of thousands of preprints posted on *biorXiv*, *medrXiv*, and *Springer-Nature* in the first six months of the pandemic (Muric, et al., 2020). By using state-of-the-art gender inference techniques, they identified the gender of almost half a million authors, which enabled them to quantify changes in the productivity of different genders. While the number of papers and authors grew in absolute terms during the early stages of the pandemic, the proportion of female authors is lower than expected. The data showed that COVID-19 exacerbated the gender gap in research production, and fewer women were engaged in COVID-19-related research.

Further, a study conducted in January 2021 found that while the initial effects of the pandemic on scientists' research time had improved, there was a decline in the number of new projects being started. This decline was more significant for female scientists and those with young children and was consistent across all research fields (Gao, et al., 2021).

These early studies were typically conducted within six to twelve months of the outbreak and analyzed only modest samples of researchers, leaving open questions about the duration and heterogeneity of pandemic effects. Our study extends the observation window to December 2022 and introduces three methodological refinements. First, we calculate relative deviation from field-specific, pre-pandemic baselines, enabling direct comparison across disciplines and institutions without scale bias. Second, we aggregate outputs in six-month intervals, a granularity that separates the initial shock from later adaptation phases while avoiding month-to-month noise. Third, we run stratified subgroup models (gender, career stage, field, and region) to expose intersectional patterns invisible in aggregate statistics. Coupled with a dataset of over 25 million works by more than 10 million authors at 1,000 institutions, these innovations provide a more nuanced and analytically robust portrait of the pandemic's impact on global research activity.

METHODS

In this section, we describe the data used in our study, detailing how research institutions were chosen for analysis, and how we operationalize constructs measuring scientific participation, productivity, and collaboration, and outline our method for estimating the pandemic's effects on these metrics.

Data

Our study uses the massive bibliographic dataset OpenAlex (Priem, et al., 2022), which provides information about entities such as publications, authors, institutions, venues, and the relationships between them. OpenAlex gathers data from various sources, including Microsoft Academic Graph, and is updated every month. We used the dump obtained on September 5th, 2023.

In processing the OpenAlex dataset, we applied the following filters to ensure data quality and relevance. First, we restricted the sample to works published between January 2000 and December 2022, yielding 178 million records. Next, we limited the dataset to publications affiliated with the top 1,000 most-productive institutions (ranked by pre-2020 output), reducing the corpus to approximately 30 million works. From here on, we refer to the ordering of institutions by their pre-2020 publication output as "institutional rank." Finally, we applied a DOI filter (requiring a valid, unique Digital Object Identifier) and a completion filter (non-missing author ID, institution ID, work ID, and publication year), resulting in 25 million works by 10 million unique authors, forming the comprehensive scope of our study. The analytical process involved grouping the data by high-level filters, such as continent, seniority, gender, and academic field, followed by a regression analysis within each group to identify trends and deviations. These are described in more detail below.

Constructs and Measurements

Below, we describe how we create constructs that measure research activity and collaboration.

Research Participation

Participation captures the number of scientists taking an active role at specific research institutions, which means conducting research and reporting findings in research publications. We operationalize *participation* at a given research institution in a specific year as the number of active authors at that institution in that year, i.e., researchers with at least one associated publication in that year.

Research Productivity

Productivity encompasses the number of novel and impactful scientific discoveries resulting from research. A simple proxy for the productivity of an academic institution is the *number of publications* associated with the institution. However, institutions may vary widely in size, creating a bias toward larger institutions. To address this issue, we also considered an *individual publication count*, which takes the average of individual authors' publication count within the respective year across the scope of each analysis.

Research Collaborations

We also study the collaboration networks of an institution's authors. To keep the institution as the unit of analysis, we restrict edges to co-authors who share that affiliation (internal collaborations). Burghardt et al. (2021) show a strong year-by-year correlation (≈ 0.8) between an institution's internal and total collaboration counts, so internal degree provides a practical proxy for overall connectedness (Burghardt, et al., 2021). External ties may reveal additional nuance; assessing them is left for future work.

We analyze collaborations for each of the 1,000 institutions by constructing a time series of six-month co-authorship graphs: undirected, unweighted networks whose nodes are authors and whose edges indicate co-authorship on papers published during each interval.

For each institution in each year, we use the following measures of internal collaboration: *average degree*, which captures the average number of internal collaborators of researchers at an institution; *average clustering coefficient*, which is the ratio of observed triangles to possible triangles in the internal collaboration network, serving as a measure of the prevalence of tightly-knit collaborative communities; and *average team size*, which represents the average number of institution co-authors for a publication.

Heterogeneity Factors

The response of researchers and institutions to pandemic disruptions may vary systematically based on their features. We explore several factors that could account for these diverse responses.

Institution Ranking and Geographic Location

To investigate these aspects, our analysis focuses on institutional-level data. We examine the top 1,000 institutions with the most publications through 2020. These institutions are categorized into five groups according to their publication volume (rank 1-200, 200-400, etc.). Given their global distribution, this approach facilitates an in-depth heterogeneity analysis, particularly examining how geographic region correlates with institutional responses to the pandemic. We limit geographic analysis to the level of continents, although finer-grained analysis is also feasible. The distribution of publications in the sample across continents is 32% Europe, 30% North America, 29% Asia, 4% Oceania, and 4% South America. Africa is excluded from this analysis due to the limited sample size of less than 350 thousand publications (about 1% of our filtered data).

Researcher Seniority and Gender

We examine the relationship between the seniority of researchers and their temporal participation and productivity over the past two decades. In each year, we divide active researchers into three categories based on their level of experience: junior, mid-career, and senior. Junior researchers are defined as those who are publishing within five years of their first publication. Mid-career authors are those who publish within six to ten years of their first publication, while senior authors are those publishing eleven or more years since their first publication. The distribution of publications in the sample across researcher seniority is 31% junior, 22% mid-career, and 47% senior.

We inferred author gender with the gender-guesser library (Pérez & Arcos, 2016) which assigns one of six labels: male, mostly male, female, mostly female, androgynous, or unknown, from first names. For analysis we merge the two male and two female labels into binary categories and omit the androgynous/unknown cases. Within our 10 million-author dataset, 38% are labelled male and 24% female; the remaining 38% lack a binary label.

Two caveats warrant emphasis. First, the underlying dictionaries draw heavily on Western census data, so coverage is weaker for many Asian and other non-Western names, inflating the unknown share and dampening regional gender contrasts. Second, this imperfect name-based approach is strictly binary and cannot capture non-binary identities. Even so, the sheer scale of the dataset and explicit accounting of uncertainty provide a signal strong enough to explore broad gender patterns. We therefore retain these measures as a useful lens on how the pandemic affected men and women differently within the global research community.

Field of Study

We assign each paper to one of 19 root fields by propagating its OpenAlex concept scores up the hierarchy to the root level. We then sum the propagated scores for each root concept and use an argmax to select the field with the highest total.

Modeling Pandemic Disruption Effects

To estimate the effect of the pandemic disruption, we adopted an approach used in previous studies (Muric, et al., 2020). Specifically, we performed linear regression for a construct over a time period of 10 years and used the trained model to predict the value of the construct for the next time point. We measure the deviation, i.e., the difference between the actual and predicted values of the construct. We then slide the time window over by one year and repeat the procedure. We calculate this *relative deviation*, i.e., deviation (difference between the actual and predicted value) divided by the predicted value.

We compute a construct's mean and standard deviation across all selected institutions and plot the mean for the observed groups throughout this paper. Focusing on 2020, this approach allows us to estimate the counterfactual: how the construct would have evolved if it were not for the pandemic. Then, a deviation of the real value from its predicted value gives the effect of the pandemic.

To improve temporal resolution without introducing excessive date noise, we segment the series into half-year windows: H1 (January–June) and H2 (July–December). Given field-specific submission lags and staggered regional outbreak timelines, we use January 2020 as a pragmatic, globally comparable disruption onset, marking 2020 H1 as the initial phase. Publication dates in OpenAlex are often defaulted to 1 January when the exact month is missing; bins shorter than six months would therefore amplify artefactual spikes, whereas half-year bins smooth this effect while still distinguishing the initial 2020 shock from subsequent adaptation. A global phase taxonomy (e.g., "outbreak", "lockdown", "recovery") is not applied because pandemic timelines differed markedly across regions – China's strict controls, Sweden's light-touch approach, and the United States' state-level heterogeneity overlapped imperfectly. A uniform six-month grid provides a neutral, comparable baseline; regional or field-specific phase shifts can then be examined as deviations from that baseline rather than baked into the segmentation. To correct the January-date imbalance we fit separate regressions for H1 and H2 and merge the resulting deviations, which explains the zig-zag pattern visible in some figures. Further, our interrupted-time-series approach treats early 2020 as a single, uniform "shock," but research projects progress at different paces – some experiments begun before lockdown only lead to publications months or years later. As a result, our estimates may mix immediate bursts in activity with slower, delayed effects. Future work can use models that allow multiple change points or gradual trend shifts to separate the pandemic's rapid impacts from those that unfolded more slowly.

Our time-series linear regression is estimated with ordinary least squares (OLS). We assess model fit with the *coefficient of determination (R^2),* which indicates the proportion of variance in the data explained by the model (values closer to 1 signal a better fit). For participation and productivity metrics, R^2 *scores* were consistently high, ranging from a minimum of 60% to often between 75% and 95%. However, in the field of study analysis, we observed a wider range of R^2 *scores*, varying from 25% to 80% across different fields. Collaboration metrics presented a lower goodness of fit, with R^2 *scores* spanning from 10% to 60%, though most were above 25%.

For *Individual Productivity* we show the time-normalized percentage increase, where the normalized publication count in the year 2010 is used as a baseline measure, and each successive year shows the percentage increase or decrease in the normalized publication count.

We examine heterogeneity by stratifying the data on each factor (e.g., gender, career stage, region) and running the regression separately within each subgroup—for example, estimating the pandemic effect for female authors and male authors in separate models.

RESULTS

We analyze the constructs of research activity we defined above to measure deviation from historical trends. We mark first half of 2020 as the start of the pandemic and measure changes in constructs with respect to this date.

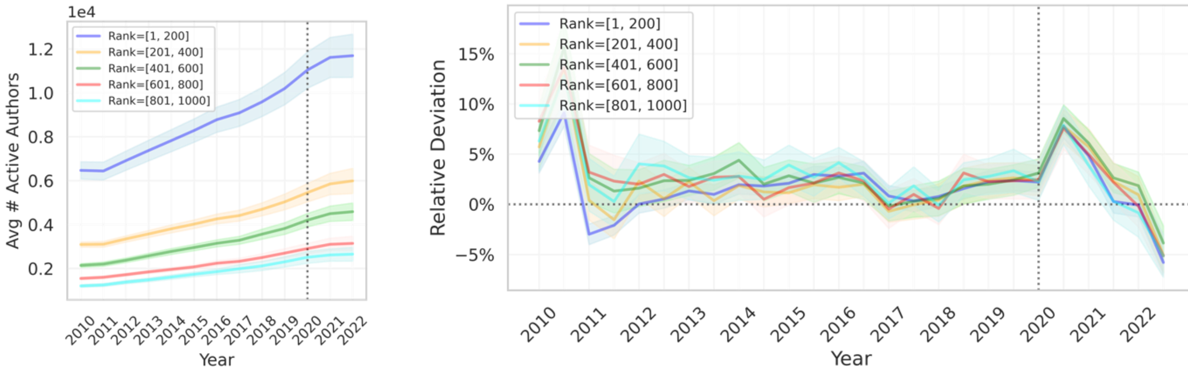
Author Participation

Figure 1 shows that researcher participation, i.e., the average number of active authors at research institutions, varies significantly across different institutional rankings (Figure 1a). However, when we plot the relative deviation from the predicted participation (Figure 1b), the effect of the institutional ranking largely disappears. Surprisingly, this is true both before the pandemic and also how the institutions are impacted by the pandemic. Specifically, across all institutional rankings, we see an increase in author participation early in the pandemic, reaching its peak (+10% relative deviation) in the second half of 2020, which then reverts to pre-pandemic trends by the end of 2022. It's important to recognize that the figures show a roughly -5% negative relative deviation in late 2022. However, this deviation is being compared to a linear model trained through the prior year, which includes the late 2020 spike. In contrast, when we examine the relative deviation versus a linear model trained on data frozen through 2019, we find that late 2022 resembles a return to pre-pandemic levels of participation. The only slight exception to this trend is among the top-ranked institutions (rank 1-200), where we observe an average decrease in author participation of - 1% in late 2022 when compared using linear regressions on the 2010-2019 data (Table 1).

Rank	2020 H1	2020 H2	2021 H1	2021 H2	2022 H1	2022 H2
1-200	0.02 ± 0.07	0.08 ± 0.09	0.07 ± 0.10	0.04 ± 0.10	0.04 ± 0.12	-0.01 ± 0.12
201-400	0.02 ± 0.11	0.08 ± 0.12	0.07 ± 0.14	0.06 ± 0.15	0.05 ± 0.19	0.01 ± 0.18
401-600	0.03 ± 0.10	0.09 ± 0.10	0.08 ± 0.12	0.07 ± 0.13	0.07 ± 0.16	0.02 ± 0.16
601-800	0.02 ± 0.12	0.07 ± 0.15	0.06 ± 0.17	0.06 ± 0.17	0.04 ± 0.21	0.00 ± 0.21
801-1000	0.02 ± 0.13	0.08 ± 0.13	0.05 ± 0.17	0.04 ± 0.16	0.02 ± 0.22	-0.01 ± 0.20

Table 1. Mean relative deviation in *Author Participation* (±95 % CI) at the 1,000 leading research institutions, by *Institutional Rank*, 2020 – 2022. Deviations are calculated from an alternative OLS baseline trained on 2010 – 2019 data; values converge toward zero by late 2022, indicating a return to pre-pandemic levels.

The R^2 values of the regression are quite high, indicating a good fit of the regression models to the data. The R^2 values depend on institution rank: they range from approximately 85% to 90% for rank 1-200 to around 70% to 75% for rank 801-1,000. Notably, higher-ranked institutions also have narrower confidence intervals, indicating that deviations within this group fall within a tighter range. The consistently positive deviations suggest super-linear growth in active authors at the top 1,000 institutions over the past two decades, indicating that a non-linear predictive model might yield more accurate predictions and higher R^2 values.



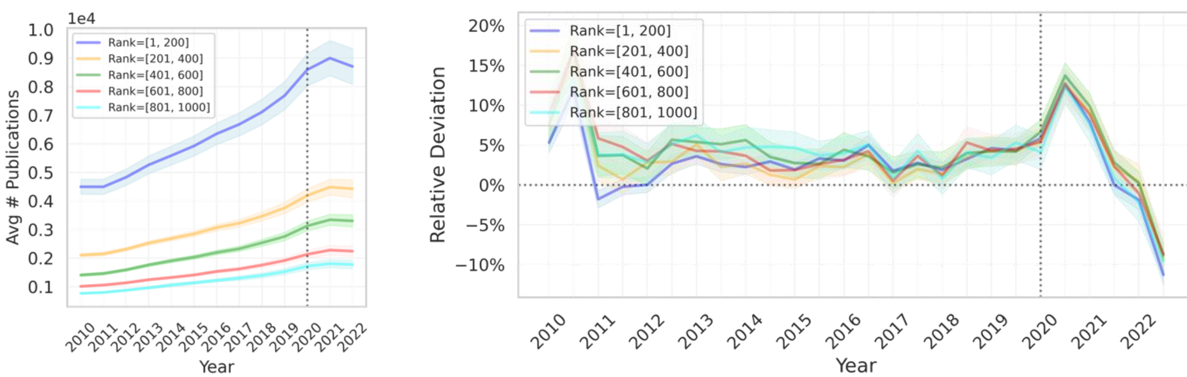
(a) **(b)**

Figure 1. *Author participation* at the 1,000 most-productive universities, by *Institutional Rank*, 2010-2022. (a) Mean active authors per institution (millions). (b) Six-month relative deviation from the preceding decade's trend. Shaded bands denote 95 % confidence intervals (CI) for each series.

Productivity

Similar observations can be extended to institutional productivity (Figure 2). Notably, the surprising surge in author participation was associated with a significant rise in the average number of publications in late 2020 and early 2021. Moreover, the observed 10% increase in the number of authors (Figure 1) was further amplified by an unexpected 15-20% boost in individual productivity (Figure 3). This dual increase culminated in a peak of 15% in institutional productivity deviation during late 2020 across all institutional rank groups in the top 1,000 institutions.

Once more, we observe a reversion to pre-pandemic total institutional publication trends by late 2022, which corresponds to the -10% trend in late 2022 depicted in Figure 2b trained on data through 2021. Notably, a similar exception is observed among the top-ranked institutions (rank 1-200), which experienced a modest decline of -4% in late 2022 when assessed through linear regressions on the 2010-2019 dataset – see Table 2 in the extended pre-print release of these results (Ahrabian, et al., 2024).



(a) **(b)**

Figure 2. Institutional Productivity at the 1,000 most-productive universities, by Institutional Rank, 2010-2022. (a) Average number of publications per institution (millions). (b) Six-month relative deviation from the preceding decade's trend. Shaded bands denote 95 % confidence intervals (CI) for each series.

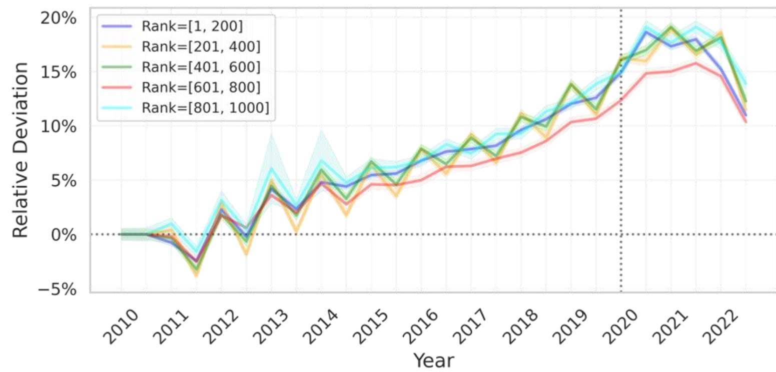
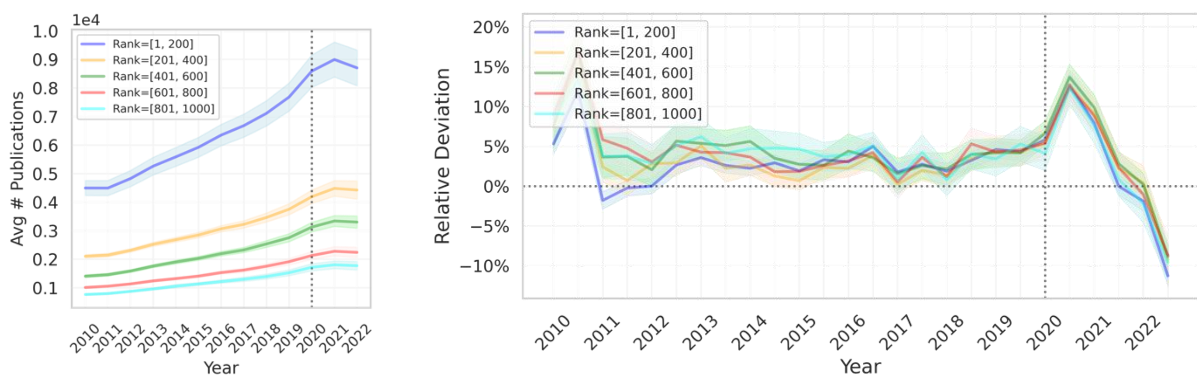


Figure 3. I*ndividual Productivity*, relative to 2010, in the top 1,000 universities grouped by *Institutional Rank*, calculated and plotted in six-month intervals between 2010-2022.

Research Collaborations

The pandemic's impact on collaborations, as shown in Figure 4 reveal intriguing dynamics (see also Figures 10 and 11 in the extended pre-print release of these results (Ahrabian, et al., 2024). The average degree, which represents the average number of internal collaborators per author, exhibits notable shifts during the pandemic. While this metric displayed a period of negative deviation over the five years prior to the pandemic, there was a noticeable rebound during the pandemic. In late 2020 and early 2021, we observed an approximate 5% increase in this metric. Remarkably, certain institutions within the top 200 ranks experienced a substantial surge of around 10%, especially within the first half of 2021 and 2022.

Surprisingly, the cohesiveness of institutional researcher networks, as illustrated by the network's clustering coefficient and the average team size, remained relatively stable in the face of the pandemic-induced shifts in work patterns. The observed deviations, while positive, were small, amounting to less than 2%.



(a) **(b)**

Figure 4. *Average number of internal collaborations* (average author degree) in the top 1,000 research institutions grouped by *institutional rank*, 2010-2022. (a) Raw data: Average number of internal collaborations per institution, grouped by rank. (b) Relative deviation for average degree, trained on previous 10 years' data, calculated and plotted in six-month intervals.

Heterogeneity Analysis

Geography

To measure how the location of an institution impacted the effects of the pandemic disruptions on research, we group institutions by continent and carry out regression analysis separately for each continent.2 Heterogeneity analysis of the top 1,000 institutions by continent shows distinct trends in Asia and South America during the pandemic. Asian institutions demonstrated remarkable resilience, with metrics like active authors (see Table 12 in the extended pre-print release of these results (Ahrabian, et al., 2024)) and publications per institution (Figure 5) aligning closely with pre-pandemic trends. This steady performance, despite the early and rigorous pandemic responses in the region, suggests a high degree of preparedness or adaptability among these institutions to the sudden shifts brought about by the pandemic.

Conversely, South American institutions appear to have faced more pronounced challenges. A -20% relative deviation in author participation and institutional productivity by late 2022 marked a notable departure from the trend, though it translates to a modest -2% change relative to the pre-pandemic norms – see Table 6 in the extended pre-print release of these results (Ahrabian, et al., 2024). This contrast highlights the impact of the pandemic on these institutions, potentially mirroring broader socio-economic and health-related struggles across the continent, which could have hindered their academic research and collaborations more significantly.

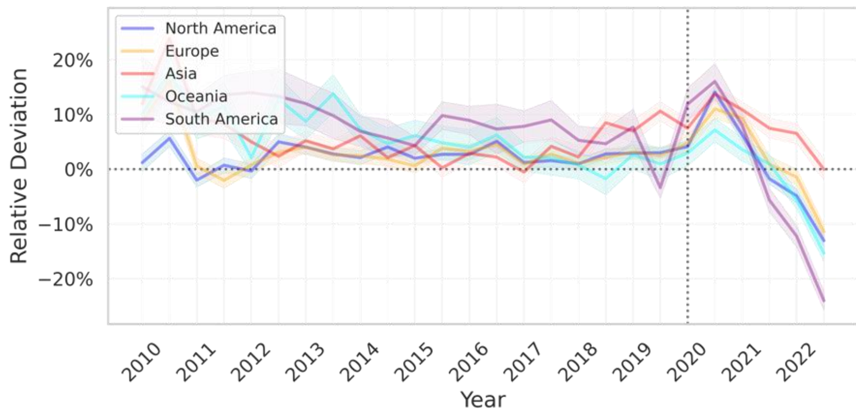


Figure 5. *Institutional Productivity* in the top 1,000 universities grouped by *Institutional Region*, 2010-2022. Relative deviation for average publications, trained on previous 10 years' data, calculated and plotted in six-month intervals.

Researcher Seniority

Figure 6 shows that early-career (junior) researchers experienced the largest surge, peaking at +12% relative deviation in H1 2020 versus +5% for mid-level and +8% for senior scholars. By late 2021, juniors still sustained +5% above baseline, compared with +1–2% for senior and near 0% for mid-level groups. Network clustering (Figure 7) followed a similar pattern: junior clustering climbed to +9 %, mid-level to +5 %, and senior to +1 % by 2022. These effect sizes (with 95% CIs consistently above zero for juniors) demonstrate a durable boost in both participation and internal network cohesion for early-career scientists. While caregiving demands may explain slower rebounds for more established scholars, future work should test how institutional support and funding mechanisms moderated these stage-specific patterns.

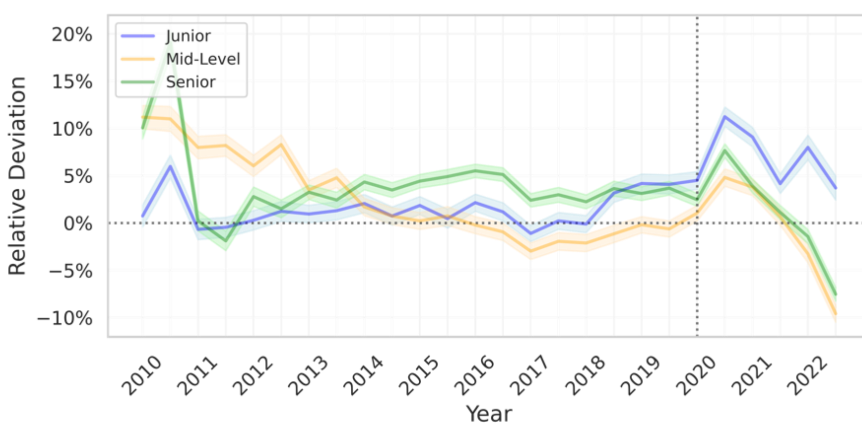


Figure 6. *Researcher Seniority*: Relative deviation of author participation over the period from 2010–2020 at the top 1,000 universities grouped by seniority level. Relative deviation for author participation, trained on previous 10 years' data, calculated and plotted in six-month intervals.

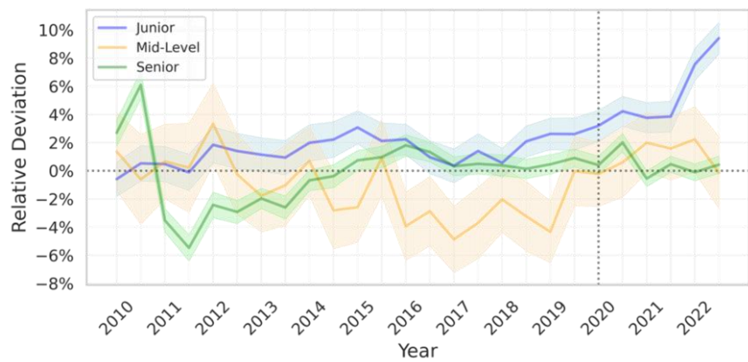


Figure 7. ***Average institutional internal network clustering coefficient*** **at the top 1,000 universities grouped by Author Seniority, 2010-2022. Relative deviation for average institutional internal network clustering coefficient, trained on previous 10 years' data, calculated and plotted in six-month intervals.**

Researcher Gender

In contrast with early-pandemic studies (Gao et al., 2021; Muric et al., 2020), our findings show that female authors marginally outpaced their peers: female participation peaked at **+10%** deviation in H2 2020 versus **+6%** for male and **+8%** for the overall cohort (Figure 8). By late 2022, female deviation receded to **–5%**, somewhat less negative than male (–8%). The non-overlapping confidence bands at the peak confirm these differences as statistically robust. Future analyses should investigate whether national family-support policies or institutional childcare provisions underpinned women's relative gains, completing the link between quantitative trends and contextual factors.

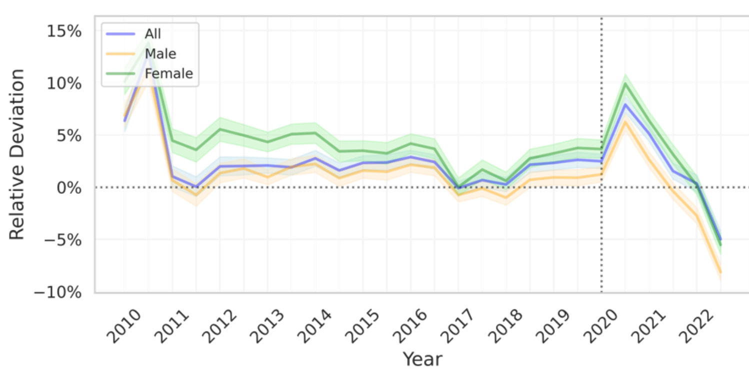


Figure 8. ***Author Participation*** **in the top 1,000 universities grouped by** ***Author Gender*****, 2010-2022. Relative deviation for author participation, trained on previous 10 years' data, calculated and plotted in six-month intervals.**

Fields of Study

In our field-specific analysis (Figure 9), Medicine drove much of the late 2020 publication surge, with a +30 % relative deviation, reflecting the urgent global response to COVID-19. In our extended pre-print release of these results we also show that other disciplines - most notably Economics and Psychology - also peaked in late 2020/early 2021 (+10%), signaling the pandemic's broad societal impacts (Ahrabian, et al., 2024).

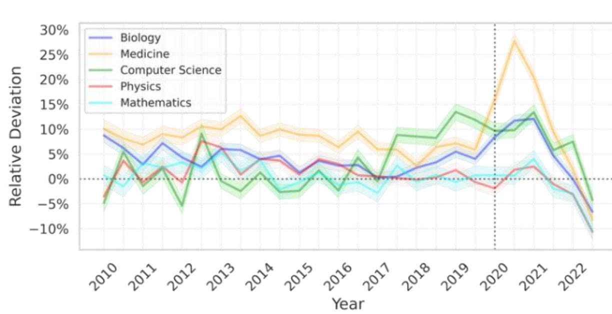


Figure 9. ***Institutional Productivity*** **in the top 1,000 universities grouped by** ***Field of Study*****, 2010-2022. Relative deviation for average publications, trained on previous 10 years' data, calculated and plotted in six-month intervals.**

CONCLUSION

Our longitudinal scientometric analysis reveals a short-lived, pandemic-induced surge in research activity that had largely dissipated by late 2022. Across the 1,000 most-productive universities, author participation and publication counts climbed ~10 % in the second half of 2020, then converged toward pre-pandemic trajectories. This pattern held regardless of institutional rank – a striking uniformity given the large baseline gaps in output and collaboration.

Regional heterogeneity, however, was pronounced. Asian institutions maintained participation and collaboration levels close to their historical trends, while South American universities experienced the sharpest downturns, suggesting (but not yet proving) that local public-health and economic conditions may have shaped recovery. Future work will test this hypothesis by adding national- and institutional-level covariates. At the individual level, early-career researchers and women registered the greatest participation gains, contradicting early, smaller-scale studies that forecasted lasting gender setbacks. These findings call for intersectional follow-up work that jointly models career stage and gender.

We hypothesise that travel bans, reduced commuting, and a shift from lab-based to desk-based tasks temporarily freed time for manuscript preparation, inflating output without altering longer-term productivity capacity. Whether the slight dip we observe in 2022 marks a simple return to baseline or the start of a deeper contraction will become clear after we incorporate the newly available 2023–2024 data into our analysis.

Our study supplies an institution-level baseline and an open analytical framework that policy makers, research managers, and fellow informetricians can use to monitor post-pandemic recovery and to stress-test research systems against future crises. Next steps include fitting non-linear models to capture super-linear growth, extending analyses to inter-institutional collaborations, and conducting intersectional and finer-grained disciplinary examinations.

LIMITATIONS AND FUTURE WORK

Our modelling strategy is deliberately streamlined and aggregate. It omits institution- and country-level covariates (e.g., funding shocks, lockdown strictness, disciplinary mix) and models each factor separately, without interaction terms; compound effects such as gender × region or field × career stage are therefore not yet captured. A next step is to add these structural variables, extend collaboration metrics to full cross-institutional networks, and explore multivariate, non-linear, and multilevel models that can reveal super-linear growth and intersectional dynamics more precisely.

GENERATIVE AI USE

We employed OpenAI ChatGPT and Anthropic Claude for the following purposes: (a) rephrasing and tightening sentences and paragraphs to improve clarity and succinctness, (b) check the spelling and grammar of the final manuscript, and (c) suggesting alternative titles, keywords, and micro-edits for style consistency. We evaluated the output by manually reviewing every suggestion for factual accuracy, alignment with our data and arguments, and appropriateness of tone, iteratively editing or discarding text as needed. The authors assume full responsibility for the content of this submission.

AUTHOR ATTRIBUTION

Rusti, Casandra: Formal analysis; Investigation; Methodology; Project administration; Validation; Visualization; Writing – original draft; Writing – review & editing**.** Ahrabian, Kian: Data curation; Formal analysis; Investigation; Methodology; Validation; Visualization; Writing – original draft; Writing – review & editing**.** Wang, Ziao: Data curation; Methodology**.** Pujara, Jay: Conceptualization; Methodology; Funding acquisition; Project administration; Supervision; Writing – review & editing**.** Lerman, Kristina: Conceptualization; Methodology; Funding acquisition; Project administration; Supervision; Writing – review & editing

ACKNOWLEDGMENTS

This work was supported by a Keston Exploratory Research Award and by Amazon. We thank the University of Southern California Center for Advanced Research Computing (CARC) for providing the computational resources used in this study. Ziao Wang also acknowledges support from the Viterbi–Tsinghua University Summer Internship Program.

REFERENCES

Ahrabian, K. et al., 2024. *Surprising Resilience of Science During a Global Pandemic: A Large-Scale Descriptive Analysis,* s.l.: arXiv preprint arXiv:2409.07710.

Burghardt, K. A., He, Z., Percus, A. G. & Lerman, K., 2020. Heterogeneous Scaling of Research Institutions. *arXiv: Physics and Society.*

Burghardt, K. A., He, Z., Percus, A. G. & Lerman, K., 2021. The emergence of heterogeneous scaling in research institutions. *Communications Physics,* IV(1), pp. 1-6.

Deville, P. et al., 2014. Career on the Move: Geography, Stratification, and Scientific Impact. *Scientific Reports,* Volume 4.

Fortunato, S. et al., 2018. Science of science. *Science,* 359(1007), p. 10.1126/science.aao0185.

Gao, J. et al., 2021. Potentially long-lasting effects of the pandemic on scientists. *Nature communications,* XII(1), pp. 1-6.

Giurge, L. M., Whillans, A. V. & Yemiscigil, A., 2021. A multicountry perspective on gender differences in time use during COVID-19. *Proceedings of the National Academy of Sciences,* 118(12).

Joseph, A., 2020. *Coronavirus pandemic forces some research labs to shut down, with uncertain futures for scientific projects.* [Online]
Available at: https://www.statnews.com/2020/03/14/coronavirus-pandemic-forces-research- labs-shut-down/

Milojevic, S., 2014. Principles of scientific research team formation and evolution. *Proceedings of the National Academy of Sciences,* 111(11), pp. 3984-3989.

Morgan, A. C. et al., 2021. The unequal impact of parenthood in academia. *Science Advances,* VII(9).

Muric, G., Lerman, K. & Ferrara, E., 2020. COVID-19 amplifies gender disparities in research. *arXiv preprint,* p. arXiv:2006.06142.

Myers, K. R. et al., 2020. Unequal effects of the COVID-19 pandemic on scientists. *Nature human behaviour,* IV(9), pp. 880-883.

Pérez, I. S. & Arcos, D., 2016. *Gender Guesser.* [Online]
Available at: https://pypi.org/project/gender-guesser/

Priem, J., Piwowar, H. & Orr, R., 2022. OpenAlex: A fully-open index of scholarly works, authors, venues, institutions, and concepts. *ArXiv,* p. https://arxiv.org/abs/2205.01833.

Raan, A. F. J. v., 2013. Universities Scale Like Cities. *PLoS ONE,* VIII(3), p. 10.1371/journal.pone.0059384.

Sinatra, R. et al., 2016. Quantifying the evolution of individual scientific impact. *Science,* 354(6312), p. https://science.sciencemag.org/content/354/6312/aaf5239.

Staniscuaski, F. et al., 2021. Gender, race and parenthood impact academic productivity during the COVID-19 pandemic: from survey to action. *Frontiers in psychology,* Volume XII.

Taylor, R. C. et al., 2019. The Scalability, Efficiency and Complexity of Universities and Colleges: A New Lens for Assessing the Higher Educational System. *ArXiv Preprint: arXiv:1910.05470.*

Vincent-Lamarre, P., Sugimoto, C. R. & Lariviere, V., 2020. Monitoring women's scholarly production during the COVID-19 pandemic. *Github Repository.*

Wang, D., Song, C. & Barabasi, A.-L., 2013. Quantifying long-term scientific impact. *Science,* 342(6154), pp. 127-132.

Way, S. F., Morgan, A. C., Larremore, D. B. & Clauset, A., 2019. Productivity, prominence, and the effects of academic environment. *Proceedings of the National Academy of Sciences,* 116(22), pp. 10729-10733.

Wuchty, S., Jones, B. F. & Uzzi, B., 2007. The Increasing Dominance of Teams in Production of Knowledge. *Science,* 316(5827), pp. 1036-1039.