

1 **Evaluating Regional Diversity in Scientific Communication: A**  
2 **Comparative Analysis of COVID-19 Preprints and Peer-Reviewed**  
3 **Publications**

4 Dong Hyun Kim<sup>a\*</sup>, Kyu Lee Jeon, B.S.<sup>b,c\*</sup>, Seng Chan You, M.D., Ph.D.<sup>b,c</sup>

5 <sup>a</sup> Yonsei University College of Medicine, Seoul, Korea

6 <sup>b</sup> Department of Biomedical Systems Informatics, Yonsei University College of Medicine,  
7 Seoul, Korea

8 <sup>c</sup> Institute for Innovation in Digital Healthcare, Yonsei University, Seoul, Korea

9 \*These authors contributed equally to this work.

10

11 **Abstract word count:** 309

12 **Manuscript word count:** 3,186

13 **Total number of figures and tables:** 5 figures (10 supplementary), 2 tables (9  
14 supplementary)

15

16 **Address for correspondence:**

17 **Seng Chan You, M.D., Ph.D.**

18 Department of Biomedical Systems Informatics, Yonsei University College of Medicine, 50-1

19 Yonsei-ro, Seodaemun-gu, 03722, Seoul, South Korea

20 Telephone: 82-2-2228-0363; Fax: 82-2-392-8133; E-mail: [chandryou@yuhs.ac](mailto:chandryou@yuhs.ac)

21

## 22 **Abstract**

23 **Background:** The unprecedented COVID-19 pandemic has triggered extensive global  
24 research, leading to an overwhelming surge in publications with surge of preprints. Despite  
25 the proliferation of preprints during the pandemic, the specific details of their implications for  
26 global diversity, along with their utility, remain underexplored. In this study, we assess the  
27 contribution of COVID-19 preprints in diverse aspects.

28 **Methods:** We collected COVID-19-related peer-reviewed papers and preprints from  
29 SCOPUS and MedRxiv, respectively, between December 2019 to November 2022. We  
30 analyzed four key aspects of scientific communication: 1) international co-authorship patterns  
31 using network analysis and eigenvector centrality, 2) publication patterns through relative  
32 ratio analysis comparing preprint to peer-reviewed paper counts, 3) social media  
33 dissemination through analysis of X (formerly Twitter) post quotations, and 4) citation  
34 impact by comparing citation counts between peer-reviewed papers with and without preprint  
35 history. All analyses were stratified by country income levels and geographical regions.

36 **Results:** Network analysis revealed higher co-authorship diversity in preprints, with Sub-  
37 Saharan Africa, Latin America, and the Caribbean showing 3.9 to 4.5 times higher  
38 eigenvector centrality compared to peer-reviewed papers. Countries with lower GDP showed  
39 significantly higher preprint publication ratios (correlation coefficient: -0.38, p-value <  
40 0.001). Social media analysis demonstrated higher engagement with preprints, as evidenced  
41 by higher median numbers of social media quotations for preprints across all income groups.  
42 Peer-reviewed papers with preprint history received significantly higher citations (median: 10,  
43 IQR: 3-30) compared to those without (median: 5, IQR: 1-15, p-value < 0.001), particularly  
44 pronounced in low- and middle-income countries.

45 **Conclusion:** This study demonstrates the significant role of preprints in advancing regional

46 diversity in scientific communication during the COVID-19 pandemic. Our findings show  
47 enhanced international collaboration through preprints, particularly benefiting researchers  
48 from lower-income regions, higher social media engagement across income groups, and  
49 increased citation impact for papers with preprint history. These results highlight preprints as  
50 an important tool for promoting more equitable global scientific discourse.

51

## 52 **Introduction**

53 The COVID-19 pandemic has triggered extensive research efforts worldwide, with a  
54 substantial body of work being produced to understand and combat the virus.[1-5] However,  
55 much of this research has been concentrated in high-income countries, leading to disparities  
56 in the geographical representation of scientific output. Despite the global nature of the  
57 pandemic, researchers from low- and middle-income countries (LMICs) have been  
58 underrepresented, further highlighting the existing inequities in the global research landscape.

59 In this context, preprints have gained attention as a potential solution to overcome the  
60 limitations of traditional peer-reviewed journals. The conventional academic publishing  
61 process often faces issues such as lengthy and complex review processes, high publication  
62 costs, and sometimes bias based on authors' affiliations or nationalities.[6] In contrast,  
63 preprints, which involve sharing preliminary manuscripts on public servers before formal  
64 peer review, offer advantages such as rapid feedback collection, idea prioritization, and  
65 improved access to research.[7, 8] The use of preprints surged during the COVID-19  
66 pandemic, particularly in medical research.[2, 9-11] The preprint system enables swift  
67 dissemination of information in emergency situations, complementing the limitations of  
68 traditional publishing systems.

69 Although numerous preprints have been published via open server during this  
70 pandemic,[12] little attention has been given to their implications for the inclusive science  
71 communication.[13] Meanwhile, social media tools are increasingly part of the research  
72 workflow, offering new dissemination and communication possibilities to researchers.[14]

73 The aim of this study is to analyze the impact of COVID-19-related preprints and  
74 peer-reviewed papers on global scientific communication. Specifically, we assess the role of  
75 preprints in fostering more inclusive research practices by examining regional diversity in

76 authorship, publication, and dissemination. Our analysis draws on data from both publication  
77 databases and social media (X, formerly Twitter) to compare the reach and engagement of  
78 preprints and peer-reviewed papers, with a focus on regional disparities.

79

## 80 **Methods**

### 81 **Data Collection**

82 This study collected data from Scopus and medRxiv to assess regional diversity in scientific  
83 communication during COVID-19. Scopus is a peer-reviewed literature database, while  
84 medRxiv is a health sciences preprint repository.[15, 16] COVID-19-related articles from  
85 December 1, 2019, to November 31, 2022, were collected using keywords such as “COVID-  
86 19”, “SARS-CoV-2”, and “coronavirus disease 2019” (**Supplementary Appendix 1**).[1] By  
87 April 20, 2023, metadata including titles, digital object identifiers (DOIs), and author  
88 affiliations were obtained, resulting in 277,366 papers from Scopus and 19,584 from  
89 medRxiv.

90 For regional analysis, the author’s affiliations were processed using the  
91 ‘hugofitipaldi/affiliation’ R package. X (formerly Twitter) data were integrated using  
92 Altmetric and the X/Twitter API, tracking posts linking to papers and categorizing users into  
93 demographic groups such as “public”, “researcher”, “practitioner”, and “science  
94 communicator”. [17] Citation data were retrieved via the CrossRef API to compare citation  
95 counts between peer-reviewed papers with and without preprint history.[18]

96

### 97 **Regional and Income-Level Categorization**

98 For the analysis of regional and income-level diversity, country information was grouped  
99 based on the World Bank’s 2022 regional and income group classifications.[19] Income  
100 groups were defined as follows: low income (gross national income [GNI] per capita of  
101 \$1,135 or less), lower-middle income (\$1,136 to \$4,465), upper-middle income (\$4,466 to  
102 \$13,845), and high income (\$13,846 or more). Geographical regions were grouped according

103 to the World Bank’s administrative classifications, including East Asia and Pacific, Europe  
104 and Central Asia, Latin America and the Caribbean, Middle East and North Africa, North  
105 America, South Asia, and Sub-Saharan Africa. To address previous findings that Europe and  
106 North America are often overrepresented in peer-reviewed journals, we distinguished Europe  
107 from Central Asia for clearer analysis.

108

### 109 **Global Co-authorship Networks**

110 To identify patterns of international collaboration between regions, the Inter-regional Co-  
111 authorship Count was defined to quantify the number of collaborative papers between  
112 researchers from different regions. This was calculated as follow:

$$113 \quad \textit{Inter-regional Co-authorship Count}_{ij} = \sum_{k=1}^N \delta_{ik} \cdot \delta_{jk}$$

114 where  $N$  represents the total number of papers, and  $\delta_{ik}$  and  $\delta_{jk}$  indicate whether paper  $k$   
115 has authors from regions  $i$  and  $j$ , respectively.

116 To evaluate the influence of each region, we used eigenvector centrality (EVC) and  
117 calculated the Co-authorship Enhancement Index (CEI), defined as:

$$CEI_r = \frac{EVC_{preprint,r}}{EVC_{peer-reviewed,r}}$$

118 This ratio indicates whether a region  $r$  was more central in the preprint network compared  
119 to the peer-reviewed network. Eigenvector centrality assigns scores to nodes (regions) based  
120 on their connections to influential nodes, weighted by normalized inter-regional co-  
121 authorship counts. This measure highlights which regions hold the most central and

122 influential positions in global research collaboration, allowing us to compare the roles of  
123 regions in preprint versus peer-reviewed networks.

124

### 125 **Publication of Research**

126 To analyze differences in publication between preprints and peer-reviewed articles across  
127 different regions, we calculated the Publication Relative Ratio (PubRR) for each country  $c$ .

$$PubRR_c = \frac{\left( \frac{\text{Number of preprints from } c}{\text{Total preprints}} \right)}{\left( \frac{\text{Number of peer reviewed papers from } c}{\text{Total peer reviewed papers}} \right)}$$

128 PubRR represents the relative ratio of a country's share of first authors preprints to its share  
129 in peer-reviewed papers. This ratio allows us to determine whether researchers in a country  
130 are more inclined to publish preprints compared to peer-reviewed articles.

131

### 132 **Social Media Dissemination of Research**

133 To evaluate how research is disseminated on social media, we analyzed X posts of preprints  
134 and peer-reviewed papers. Social Media Quotation Count quantifies the number of X posts  
135 referencing papers from a country. A higher quotation count indicates broader dissemination  
136 of that country's research on X.

$$Social\ Media\ Quotation\ Count_c = \sum_{p \in P_c} \text{Number of X posts citing paper } p$$

137 where  $P_c$  represents the set of papers with first authors from country  $c$ .

138 Social Media Quotation Relative Ratio (QuoRR) compares the share of social media  
139 quotation count for preprints versus peer-reviewed articles for each country.

$$QuoRR_c = \frac{\left( \frac{\text{Social media quotations of preprints from } c}{\text{Total social media quotations of all preprints}} \right)}{\left( \frac{\text{Social media quotations of peer reviewed papers from } c}{\text{Total social media quotations of all peer reviewed papers}} \right)}$$



140

## 141 **Social Media Readership Patterns**

142 To understand how social media users engage with research, we used the Social Media  
143 Readership Relative Ratio (ReadRR). ReadRR measures the relative ratio of social media  
144 users citing preprints versus peer-reviewed articles from a particular country

$$ReadRR_c = \frac{\left( \frac{\text{Social media quotations from } c \text{ citing preprints}}{\text{Total social media quotations of all preprints}} \right)}{\left( \frac{\text{Social media quotations from } c \text{ citing peer reviewed papers}}{\text{Total social media quotations of all peer reviewed papers}} \right)}$$

145 To explore how the general public engages with research output in each country  
146 through social media, we calculated the Public Engagement Relative Ratio (PERR). PERR  
147 was used to assess whether public users engage more with preprints than peer-reviewed  
148 papers from a given country.

$$PERR_c = \frac{\left( \frac{\text{Public } X \text{ users' social media quotations citing preprints from } c}{\text{Total social media quotations citing preprints from } c} \right)}{\left( \frac{\text{Public } X \text{ users' social media quotations citing peer reviewed papers from } c}{\text{Total social media quotations citing peer reviewed papers from } c} \right)}$$

149

## 150 **Relative Ratio Interpretation**

151 All relative ratios (PubRR, QuoRR, ReadRR, and PERR) are interpreted as follows:

- 152     ▪ *Value > 1: greater engagement for preprints compared to peer-reviewed papers.*
- 153     ▪ *Value = 1: equal engagement.*
- 154     ▪ *Value < 1: lower engagement for preprints compared to peer-reviewed papers.*

155 These metrics provide a comprehensive understanding of collaboration, publication, social  
156 media dissemination and user engagement across different regions during the COVID-19  
157 pandemic.

158

## 159 **Statistical Analyses**

160 The coefficient of variation (CV) for inter-regional co-authorship counts was calculated to  
161 assess variability in collaboration among regions. Pearson's correlation analyses were  
162 performed to examine the relationships between GDP and four relative ratio metrics. Scatter  
163 plots and R-squared values were used to visualize these relationships. The Wilcoxon signed-  
164 rank test was used to compare the median number of social media posts quoting preprints  
165 versus peer-reviewed papers within income groups. A two-sample proportion test was  
166 employed to compare the ratio of social media quotation counts to total paper counts across  
167 Scopus and medRxiv. Chi-squared tests were used to compare user demographics citing  
168 different publication types, and the Wilcoxon rank-sum test assessed citation differences  
169 between peer-reviewed papers with and without preprint versions.

170

## 171 **Results**

### 172 **Data Collection**

173 We collected 15,413,656 X posts referencing 277,366 papers from Scopus and 2,718,034 X  
174 posts referencing 19,584 papers from medRxiv, authored by users from 201 and 213  
175 countries, respectively. For regional diversity analysis, we categorized countries of first  
176 authors and X users by income level and geographical region, excluding entries without  
177 identifiable country information (**Supplementary File 1**). The final categorization, which is  
178 shown in **Supplementary Figure 1**, included:

- 179 1. Country-classified Papers: We categorized papers based on the country of the  
180 first author for 250,442 Scopus papers (192 countries) and 13,952 medRxiv  
181 papers (147 countries) (**Supplementary Table 1**).
- 182 2. Country-classified X Posts: We categorized X posts based on the country of the  
183 paper's first author (14,325,011 posts for Scopus and 1,853,738 for medRxiv)  
184 (**Supplementary Table 2 A**). Additionally, we categorized posts based on the  
185 users' country (5,910,994 posts for Scopus and 1,010,582 for medRxiv)  
186 (**Supplementary Table 2 B**).

187

### 188 **Global Co-authorship Networks**

189 Inter-regional co-authorship counts were calculated across 8 regions, resulting in 28 pairwise  
190 combinations for each data source. The average co-authorship count was 196,572.30 for  
191 Scopus and 6,833.64 for medRxiv, with a coefficient of variation (CV) of 2.70 for Scopus and  
192 1.95 for medRxiv. This indicates a wider disparity in regional collaboration for peer-reviewed  
193 articles compared to preprints (**Table 1**). Network graphs visualizing these counts are shown

194 in **Figure 1**, with detailed matrix and counts provided in **Supplementary Figure 2** and  
195 **Supplementary Table 3**. The co-authorship enhancement index (CEI)—calculated as the  
196 ratio of EVC for preprints to peer-reviewed publications—was highest in North America  
197 (7.66), followed by Sub-Saharan Africa (4.53), Central Asia (4.29), Latin America &  
198 Caribbean (3.90), Middle East & North Africa (2.19), East Asia & Pacific (1.76), and South  
199 Asia (1.07) (**Table 2**). **Supplementary File 2** contains inter-country co-authorship networks  
200 for peer-reviewed papers and preprints, while **Supplementary File 3** provides interactive  
201 visualizations for scatter plots, geographic heatmaps, and bar plots.

202

### 203 **Publication of Research**

204 The Publication Relative Ratio (PubRR) was calculated to compare the number of papers  
205 authored by researchers from each country across Scopus and medRxiv. A significant  
206 negative correlation between PubRR and GDP was observed (Pearson correlation coefficient:  
207 -0.38, p-value < 0.001), indicating that researchers from countries with lower GDPs are more  
208 likely to publish in preprint repositories (**Figure 2 A**). Over half of the countries in North  
209 America, Latin America, Central Asia, and Sub-Saharan Africa displayed PubRR values  
210 greater than 1 (**Figure 2 B, Supplementary Figure 3**). High PubRR values were recorded in  
211 countries such as Turkmenistan (35.9), South Sudan (33.3), Mali (15.9), Cameroon (11.2) and  
212 Ukraine (9.9) (**Supplementary Table 4**). Preprints surged rapidly during COVID-19, reached  
213 half-publications faster, and had a 196 day lead over corresponding peer-reviewed  
214 publications (**Supplementary Figure 4**).

215

### 216 **Social Media Dissemination of Research**

217 We analyzed the distribution of posts on platform X that referenced preprints and peer-  
218 reviewed papers by first authors from each country. Across all income groups, the median  
219 Social Media Quotation Count for medRxiv was significantly higher than for Scopus  
220 (Wilcoxon test,  $p$ -value  $< 0.001$ ) (**Figure 3 A, Supplementary Table 5**). A significant  
221 negative correlation was found between Social Media Quotation Relative Ratio (QuoRR) and  
222 GDP (Pearson correlation coefficient:  $-0.17$ ,  $p$ -value =  $0.046$ ), indicating that countries with  
223 lower GDPs had a higher relative share of social media quotations for preprints (**Figure 3 B**).  
224 Approximately half of the countries in North America, Sub-Saharan Africa, and North-  
225 Middle Africa had QuoRR values greater than 1, suggesting relatively higher social media  
226 attention for preprints in these regions (**Figure 3 C, Supplementary Figure 5**). High QuoRR  
227 values were observed in countries like Cameroon (80.8), South Sudan (62.9), Sudan (36.7),  
228 Turkmenistan (19.3), and Kazakhstan (13.1) (**Supplementary Table 6**).

229

### 230 **Social Media Readership Patterns**

231 We compared social media engagement for preprints versus peer-reviewed papers across  
232 countries. No significant correlation was found between GDP and the Social Media  
233 Readership Relative Ratio (ReadRR) (Pearson correlation coefficient:  $-0.11$ ,  $p$ -value =  $0.136$ )  
234 (**Figure 4, Supplementary Figure 6**). Both types of papers show a similar decline in the  
235 social media quotation-to-paper ratio from high to low-income groups, with no significant  
236 difference between paper types (**Supplementary Figure 7**). The general public was the most  
237 frequent group citing both preprints and peer-reviewed papers, with a higher engagement rate  
238 for preprints (preprints:  $88.26\%$ , peer-reviewed papers:  $85.75\%$ ,  $p$ -value  $< 0.001$ )  
239 (**Supplementary Figure 8**). Scientists, practitioners and science communicators showed  
240 greater engagement with peer-reviewed articles ( $p$ -value  $< 0.001$ ). The Public Engagement

241 Relative Ratio (PERR) exhibited a negative correlation with GDP (Pearson correlation  
242 coefficient: -0.20, p-value = 0.021). In North America and Central Asia, all countries had  
243 PERR values greater than 1. Similarly, in East Asia & Pacific, Sub-Saharan Africa, and Latin  
244 America & Caribbean, approximately 90% of countries had PERR greater than 1  
245 (**Supplementary Figure 9**). High PERR values were observed in countries like Algeria (1.2),  
246 Palestine (4.4), Cameroon (1.3), Kazakhstan (1.2), Armenia (1.2), and Cambodia (1.2)  
247 (**Supplementary Table 7**).

248

## 249 **Citation Analyses**

250 We assessed the impact of prior preprint publication on subsequent citation counts in peer-  
251 reviewed journals. Papers with a preprint history had significantly higher citation counts  
252 (median: 10, Q1–Q3: 3–30) compared to those without (median: 5, Q1–Q3: 1–15) (p-value <  
253 0.001) (**Figure 5**). This trend held true for papers from low- and middle-income countries  
254 (LMICs), with preprints receiving higher citation counts (median: 9, Q1- Q3: 2-25) than  
255 those without (median: 2, Q1-Q3: 0-8) (p-value < 0.001). The median citation counts by  
256 country for preprints and peer-reviewed papers are shown in **Supplementary Figure 10** and  
257 citation count statistics by region and income group for peer-reviewed papers and preprints in  
258 **Supplementary Table 8**. We also tracked changes in the income classification of first authors  
259 when preprints transitioned to peer-reviewed publications. Of these, 57% of preprints from  
260 low-income countries, 49% from lower-middle-income countries, and 22% from upper-  
261 middle-income countries were subsequently published with first authors from higher-income  
262 groups. Among LMIC preprints, 22 remained published by LMIC authors, while 26  
263 transitioned to authorship from high-income countries (**Supplementary Table 9**).

264

## 265 **Discussion**

266 This study provides a comprehensive bibliometric analysis of how preprints influenced  
267 regional diversity in scientific communication during the COVID-19 pandemic. Our  
268 comparative analysis of preprints and peer-reviewed publications revealed distinct patterns  
269 across geographical regions and economic strata. Three key findings emerged: First, preprints  
270 significantly enhanced international collaboration by facilitating the inclusion of authors from  
271 regions traditionally underrepresented in peer-reviewed publications. Second, researchers  
272 from lower-GDP countries showed higher engagement with preprint platforms, suggesting  
273 these platforms serve as a more accessible publication route. Third, preprints demonstrated  
274 broader reach through social media engagement and, notably, papers that began as preprints  
275 achieved higher citation counts after peer review, particularly those from lower-GDP regions.  
276 These findings collectively indicate that preprints serve as an effective mechanism for  
277 democratizing scientific communication, enabling researchers from underrepresented regions  
278 to increase their visibility and impact in the global scientific discourse.

279 Our study found that peer-reviewed journals tend to publish papers from established  
280 collaborative networks, especially in affluent regions such as North America, Europe, and  
281 East Asia. This pattern likely represents several factors including long-standing research  
282 relationships,[20-22] greater access to funding,[23, 24] and the prestige of established  
283 institutions.[25-27] In contrast, our analysis revealed that preprints demonstrated a more  
284 balanced distribution of collaborations across regions, with particularly strong representation  
285 from lower-GDP countries. The higher utilization of preprint servers by researchers from  
286 lower-GDP countries compared to traditional peer-reviewed journals suggests that preprints  
287 provide a more accessible publication route for these researchers.

288 This accessibility proved especially crucial during the COVID-19 pandemic,[28] where rapid

289 dissemination of research findings was essential for global public health responses.[29] By  
290 enabling researchers from lower-income regions to contribute more readily to the scientific  
291 discourse, preprints helped foster greater inclusivity in scientific communication during a  
292 critical period.

293 Our analysis further revealed that preprints were disseminated more extensively on  
294 social media, particularly by public users. This trend was especially pronounced among  
295 papers by researchers from lower-GDP countries, for whom social media served as a crucial  
296 tool to enhance visibility and facilitate engagement with a broader audience.[30, 31] The role  
297 of social media in disseminating research is critical, as it provides an alternative avenue for  
298 researchers who may lack the traditional means of publication and collaboration.[30, 32]  
299 Previous studies have shown that social media can significantly increase the citation impact  
300 of research, particularly for authors from institutions with fewer resources.[33, 34] This  
301 suggests that the use of social media not only facilitates the dissemination of research but also  
302 democratizes access to scientific knowledge, allowing researchers from lower GDP countries  
303 to engage more actively in global discourse.

304 Our analyses also indicate that peer-reviewed papers that originated as preprints tend  
305 to receive a significantly higher median citation count compared to those that did not. This  
306 trend is especially pronounced for papers where LMIC researchers are the first authors. The  
307 increased visibility and dissemination afforded by preprints may enhance subsequent citations,  
308 as preprints often serve as a preliminary platform for sharing research findings, which can  
309 lead to greater engagement and discourse within the academic community.[35-37] However,  
310 it has been observed that LMIC authorship in preprints often shifts to higher-income  
311 countries in peer-reviewed versions, likely due to disparities in funding, institutional support,  
312 and perceived credibility.[23, 38] This raises concerns about equity and representation in



313 global research outputs.[39, 40]

314 Despite the advantages of preprints, concerns regarding their quality and credibility  
315 persist due to the absence of formal peer review prior to publication. Addressing these  
316 concerns requires fostering an open discussion environment with interactive feedback  
317 mechanisms.[41] Utilizing communication tools such as forums and collaborative platforms  
318 can enhance the quality assurance of preprints by allowing researchers to engage in  
319 constructive dialogue about the findings presented.[42] This approach is supported by the  
320 notion that preprints can be revised based on community input. Furthermore, the integration  
321 of communication tools such as social media can facilitate these discussions, allowing  
322 researchers to share insights and critiques in real-time. The use of interactive feedback  
323 mechanisms can help establish frameworks, allowing for a more robust evaluation of  
324 preprints. This aligns with the findings of Weissgerber et al., who noted that the  
325 overwhelming influx of preprints during the pandemic necessitated better monitoring and  
326 assessment strategies.[43] By leveraging communication tools, researchers can create a  
327 collaborative environment that not only addresses quality concerns but also fosters a culture  
328 of transparency and accountability in scientific communication.

329 Although this study provides valuable insights, there are some limitations. One  
330 limitation relates to dataset attrition bias. A total of 9.7% of peer-reviewed papers and 28.8%  
331 of preprints were excluded due to incomplete affiliation details, despite the use of advanced  
332 methods to extract country information. This higher exclusion rate for preprints is largely due  
333 to the preprint system's reliance on free-format entries for affiliations, lacking the  
334 standardized format used in peer-reviewed papers, which complicates accurate geographic  
335 extraction. Additionally, attempts to automate web-API searches were hindered by  
336 inconsistent author identification. The reliance on SCOPUS for peer-reviewed papers and

337 MedRxiv for preprints may have excluded relevant publications from other platforms.[44]  
338 English-language queries may have missed non-English titles, particularly early COVID-19  
339 studies from China.[45] Furthermore, the country data were based on first authors' affiliations  
340 rather than their origin, and this study did not analyze time trends or journal impact factors,  
341 which could further refine the findings. Future research could address these aspects.

342

### 343 **Conclusion**

344 The COVID-19 pandemic has clearly demonstrated how inequalities in scholarly  
345 communication can be exacerbated during a global health crisis. While it is unlikely that  
346 preprints will fully replace peer-reviewed journals in the medical field, they have proven to  
347 be an important tool in mitigating these inequalities. With specific and measurable objectives,  
348 collaborative efforts can contribute to the development of an inclusive and sustainable  
349 scientific communication environment.

350

351 **Acknowledgement**

352 We extend our sincere gratitude to Dr. Joshua Wallach from Emory University for his  
353 invaluable insights and constructive feedback that significantly enhanced this meta-research  
354 study.

355

## 356 References

- 357 1. Ioannidis, J.P.A., et al., *The rapid, massive growth of COVID-19 authors in the scientific literature*.  
358 Royal Society Open Science, 2021. **8**(9): p. 210389.
- 359 2. Else, H., *How a torrent of COVID science changed research publishing - in seven charts*. Nature, 2020.  
360 **588**(7839): p. 553.
- 361 3. Majumder, M.S. and K.D. Mandl, *Early in the epidemic: impact of preprints on global discourse about*  
362 *COVID-19 transmissibility*. The Lancet Global Health, 2020. **8**(5): p. e627-e630.
- 363 4. Stoye, E., *China coronavirus: how many papers have been published?* Nature, 2020.
- 364 5. Kana, M.A., R. LaPorte, and A. Jaye, *Africa's contribution to the science of the COVID-19/SARS-CoV-*  
365 *2 pandemic*. BMJ Global Health, 2021. **6**(3): p. e004059.
- 366 6. Krumholz, H.M., *The End of Journals*. Circulation. Cardiovascular Quality and Outcomes, 2015. **8**(6):  
367 p. 533-534.
- 368 7. Bagdasarian, N., G.B. Cross, and D. Fisher, *Rapid publications risk the integrity of science in the era of*  
369 *COVID-19*. BMC Medicine, 2020. **18**(1): p. 192.
- 370 8. Ross, H.M.K., Theodora Bloom, Joseph S., *Preprints can fill a void in times of rapidly changing*  
371 *science*, in *STAT*. 2020.
- 372 9. Flanagan, A., P.B. Fontanarosa, and H. Bauchner, *Preprints Involving Medical Research—Do the*  
373 *Benefits Outweigh the Challenges?* JAMA, 2020. **324**(18): p. 1840-1843.
- 374 10. Fraser, N., et al., *Preprinting the COVID-19 pandemic*. 2020, Scientific Communication and Education.
- 375 11. Rubin, E.J., et al., *Medical Journals and the 2019-nCoV Outbreak*. The New England Journal of  
376 Medicine, 2020. **382**(9): p. 866.
- 377 12. Iyandemye, J. and M.P. Thomas, *Low income countries have the highest percentages of open access*  
378 *publication: A systematic computational analysis of the biomedical literature*. PLOS ONE, 2019. **14**(7):  
379 p. e0220229.
- 380 13. Priem, J. and K.L. Costello, *How and why scholars cite on Twitter*. Proceedings of the American  
381 Society for Information Science and Technology, 2010. **47**(1): p. 1-4.
- 382 14. Ferreira, M.R., P. Mongeon, and R. Costas, *Large-Scale Comparison of Authorship, Citations, and*  
383 *Tweets of Web of Science Authors*. Journal of Altmetrics, 2021. **4**(1): p. 1.
- 384 15. Falagas, M.E., et al., *Comparison of PubMed, Scopus, Web of Science, and Google Scholar: strengths*  
385 *and weaknesses*. FASEB journal: official publication of the Federation of American Societies for  
386 Experimental Biology, 2008. **22**(2): p. 338-342.
- 387 16. Else, H., *How a torrent of COVID science changed research publishing — in seven charts*. Nature,  
388 2020. **588**(7839): p. 553-553.
- 389 17. *Homepage*. Altmetric.
- 390 18. *REST API - Crossref*.
- 391 19. *World Bank Country and Lending Groups – World Bank Data Help Desk*. Available from:  
392 [https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups)  
393 [groups](https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups).
- 394 20. Bu, Y., et al., *Understanding persistent scientific collaboration*. Journal of the Association for  
395 Information Science and Technology, 2018. **69**(3): p. 438-448.
- 396 21. Chinchilla-Rodríguez, Z., et al., *Blockmodeling of co-authorship networks in library and information*  
397 *science in Argentina: a case study*. Scientometrics, 2012. **93**(3): p. 699-717.
- 398 22. Guerrero Bote, V.P., C. Olmeda-Gómez, and F. De Moya-Anegón, *Quantifying the benefits of*  
399 *international scientific collaboration*. Journal of the American Society for Information Science and  
400 Technology, 2013. **64**(2): p. 392-404.
- 401 23. Smith, M.J., et al., *The Scientific Impact of Nations: Journal Placement and Citation Performance*.  
402 PLOS ONE, 2014. **9**(10): p. e109195.
- 403 24. Tubbs-Cooley, H.L., et al., *Development of a Regional Nursing Research Partnership for Academic*  
404 *and Practice Collaborations*. Nursing Research and Practice, 2013. **2013**(1): p. 473864.
- 405 25. Chinchilla-Rodríguez, Z., C.R. Sugimoto, and V. Larivière, *Follow the leader: On the relationship*  
406 *between leadership and scholarly impact in international collaborations*. PLOS ONE, 2019. **14**(6): p.  
407 e0218309.
- 408 26. Onyancha, O.B., *Regional and international research collaboration and citation impact in selected*  
409 *sub-Saharan African countries in the period 2000 to 2019*. Global Knowledge, Memory and  
410 Communication, 2020. **70**(6/7): p. 577-594.
- 411 27. Limayanta, C.H., et al., *Relationship between collaboration and normalized scientific impact in*  
412 *South American public universities*. Scientometrics, 2022. **127**(11): p. 6391-6411.
- 413 28. Gianola, S., et al., *Characteristics of academic publications, preprints, and registered clinical trials on*

- 414 *the COVID-19 pandemic*. PLOS ONE, 2020. **15**(10): p. e0240123.
- 415 29. Teixeira da Silva, J.A., H. Bornemann-Cimenti, and P. Tsigaris, *Optimizing peer review to minimize the*  
416 *risk of retracting COVID-19-related literature*. *Medicine, Health Care and Philosophy*, 2021. **24**(1): p.  
417 21-26.
- 418 30. Giustini, D.M., et al., *Effective uses of social media in public health and medicine: a systematic review*  
419 *of systematic reviews*. *Online Journal of Public Health Informatics*, 2018. **10**(2): p. e62121.
- 420 31. Gates, A., et al., *Dissemination of evidence in paediatric emergency medicine: a quantitative*  
421 *descriptive evaluation of a 16-week social media promotion*. *BMJ Open*, 2018. **8**.
- 422 32. Tunnecliff, J., et al., *The Acceptability Among Health Researchers and Clinicians of Social Media to*  
423 *Translate Research Evidence to Clinical Practice: Mixed-Methods Survey and Interview Study*. *Journal*  
424 *of Medical Internet Research*, 2015. **17**(5): p. e4347.
- 425 33. Rassi, R.E., et al., *Medical research productivity in the Arab countries: 2007-2016 bibliometric*  
426 *analysis*. *Journal of Global Health*, 2018. **8**(2): p. 020411.
- 427 34. Dol, J., et al., *Health Researchers' Use of Social Media: Scoping Review*. *Journal of Medical Internet*  
428 *Research*, 2019. **21**(11): p. e13687.
- 429 35. Bertin, M. and I. Atanassova, *Preprint citation practice in PLOS*. *Scientometrics*, 2022. **127**(12): p.  
430 6895-6912.
- 431 36. Traag, V.A., *Inferring the causal effect of journals on citations*. *Quantitative Science Studies*, 2021.  
432 **2**(2): p. 496-504.
- 433 37. Subramanian, K., et al., *Pharmaceutical industry–authored preprints: scientific and social media*  
434 *impact*. *Current Medical Research and Opinion*, 2021. **37**(2): p. 269-273.
- 435 38. Eldridge, L., et al., *Authorship of Publications Supported by NCI-Funded Grants Involving Low- and*  
436 *Middle-Income Countries*. *JAMA Network Open*, 2024. **7**(3): p. e243215.
- 437 39. Sofi-Mahmudi, A., et al., *Geographical and Gender Diversity in Cochrane and non-Cochrane Reviews*  
438 *Authorship: A Meta-Research Study*. 2024.
- 439 40. Schneider, H. and N. Maleka, *Patterns of authorship on community health workers in low-and-middle-*  
440 *income countries: an analysis of publications (2012–2016)*. *BMJ Global Health*, 2018. **3**(3): p.  
441 e000797.
- 442 41. Blatch-Jones, A., A.R. Saucedo, and B. Giddins, *The use and acceptability of preprints in health and*  
443 *social care settings: a scoping review*. 2023.
- 444 42. Zeng, L., *Changes in health communication in the age of COVID-19: A study on the dissemination of*  
445 *preprints to the public*. *Frontiers in Public Health*, 2023. **11**.
- 446 43. Weissgerber, T., et al., *Automated screening of COVID-19 preprints: can we help authors to improve*  
447 *transparency and reproducibility?* *Nature Medicine*, 2021. **27**(1): p. 6-7.
- 448 44. Irawan, D.E., H. Zahroh, and I. Puebla, *Preprints as a driver of open science: Opportunities for*  
449 *Southeast Asia*. *Frontiers in Research Metrics and Analytics*, 2022. **7**.
- 450 45. Fidahic, M., et al., *Research methodology and characteristics of journal articles with original data,*  
451 *preprint articles and registered clinical trial protocols about COVID-19*. *BMC medical research*  
452 *methodology*, 2020. **20**(1): p. 161.

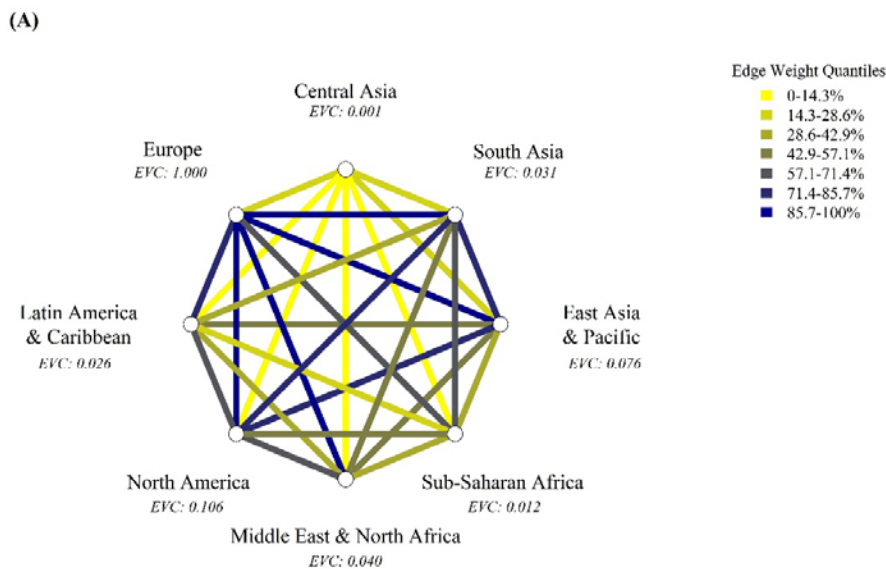
453

454

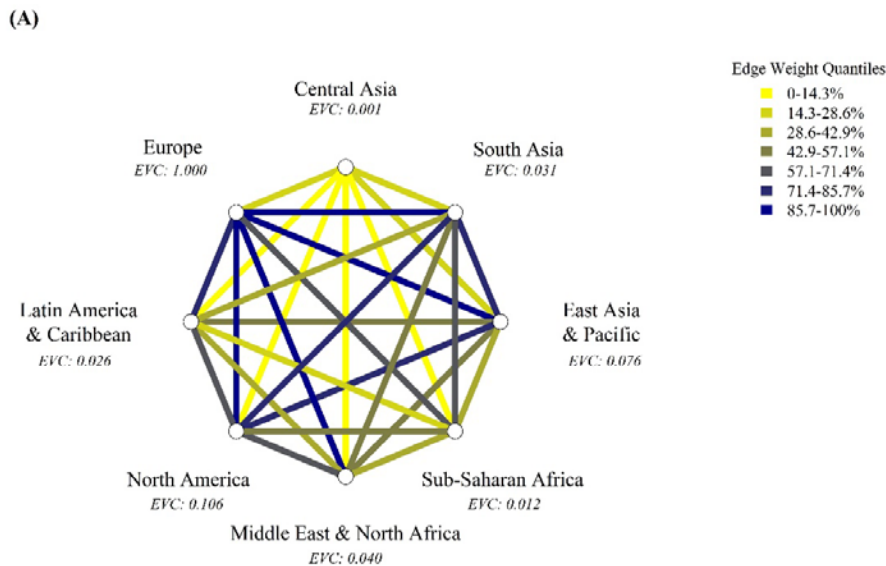
455 **Figure Legends**

456 **Figure 1. Network visualization on inter-regional co-authorship**

457 **(A) Peer-reviewed Papers (B) Preprints**



458



459

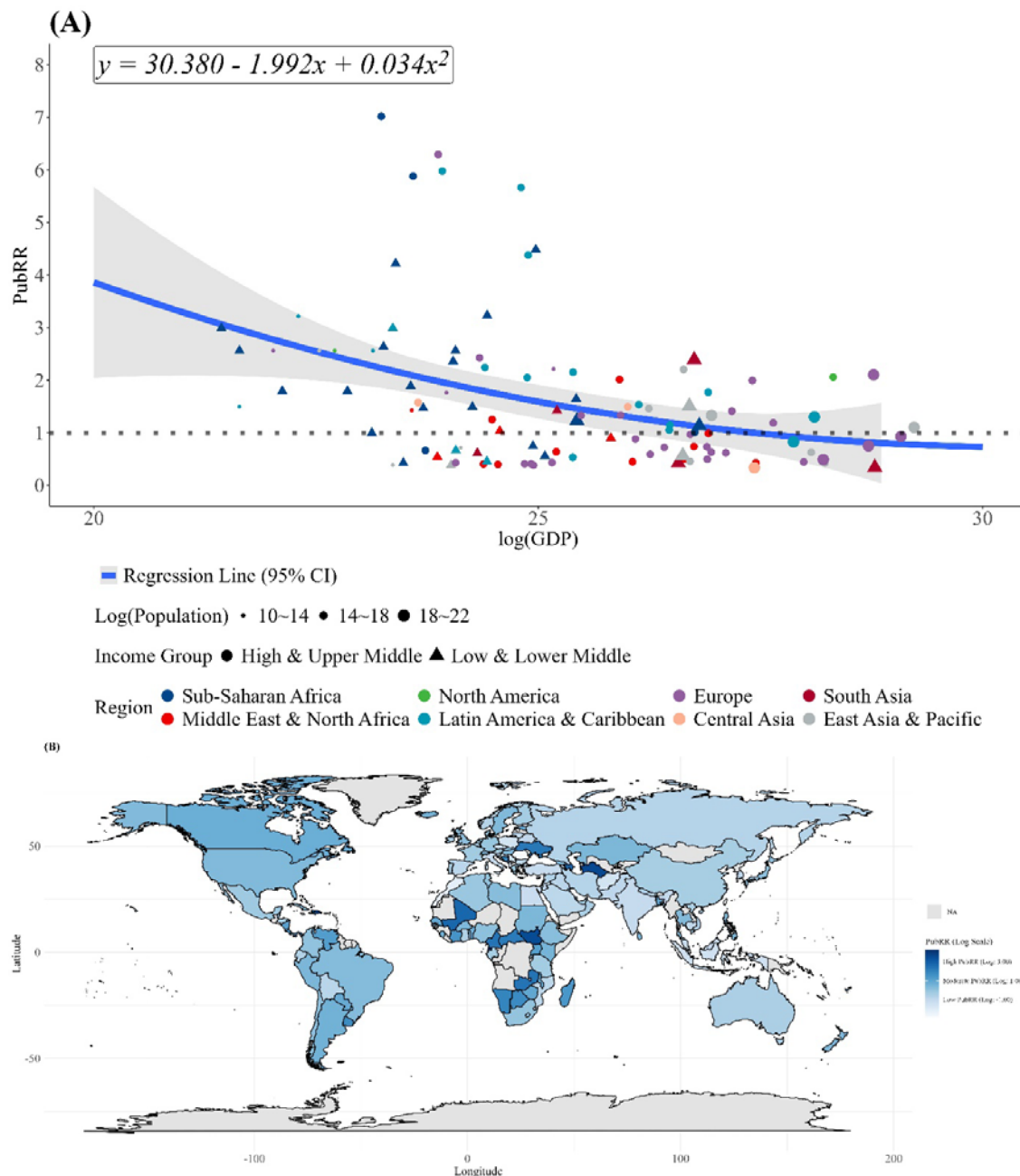
460 Figure 1, which is a network visualization on inter-regional co-authorship among regions for (a) Peer-reviewed papers and (b) Preprints, depicts regions as nodes and established connections as edges. It was made by first  
461 extracting regions for each distinct institution when papers had multiple affiliations then forming inter-regional  
462 co-authorship connections between them. The color of the edges between nodes indicates the strength of  
463 collaboration between two regions, categorized into six quantile levels. The first quantile (0-16.7) shows yellow  
464

465 for the lowest strength and last quantile (83.3-100) shows darkblue for the highest collaboration strength. The  
466 quantile cutoff values for each type are as follows: for peer-reviewed papers, the quantile cutoffs are 2,725.2,  
467 33,622.9, 55,040.8, 108,458.0, and 221,283.1; for preprints, the cutoffs are 181.4, 960.8, 2,186.1, 4,900.5, and  
468 6,888.6. The CV, EVC, and CEI was calculated from this network.  
469 **Abbreviations:** CV, Coefficient of Variation; EVC, Eigenvector Centrality; CEI, Co-authorship Enhancement  
470 Index.

471

472

473 **Figure 2. (A) Scatter plot of the country's GDP and PubRR, (B) Geographic heatmap of PubRR**



474

475

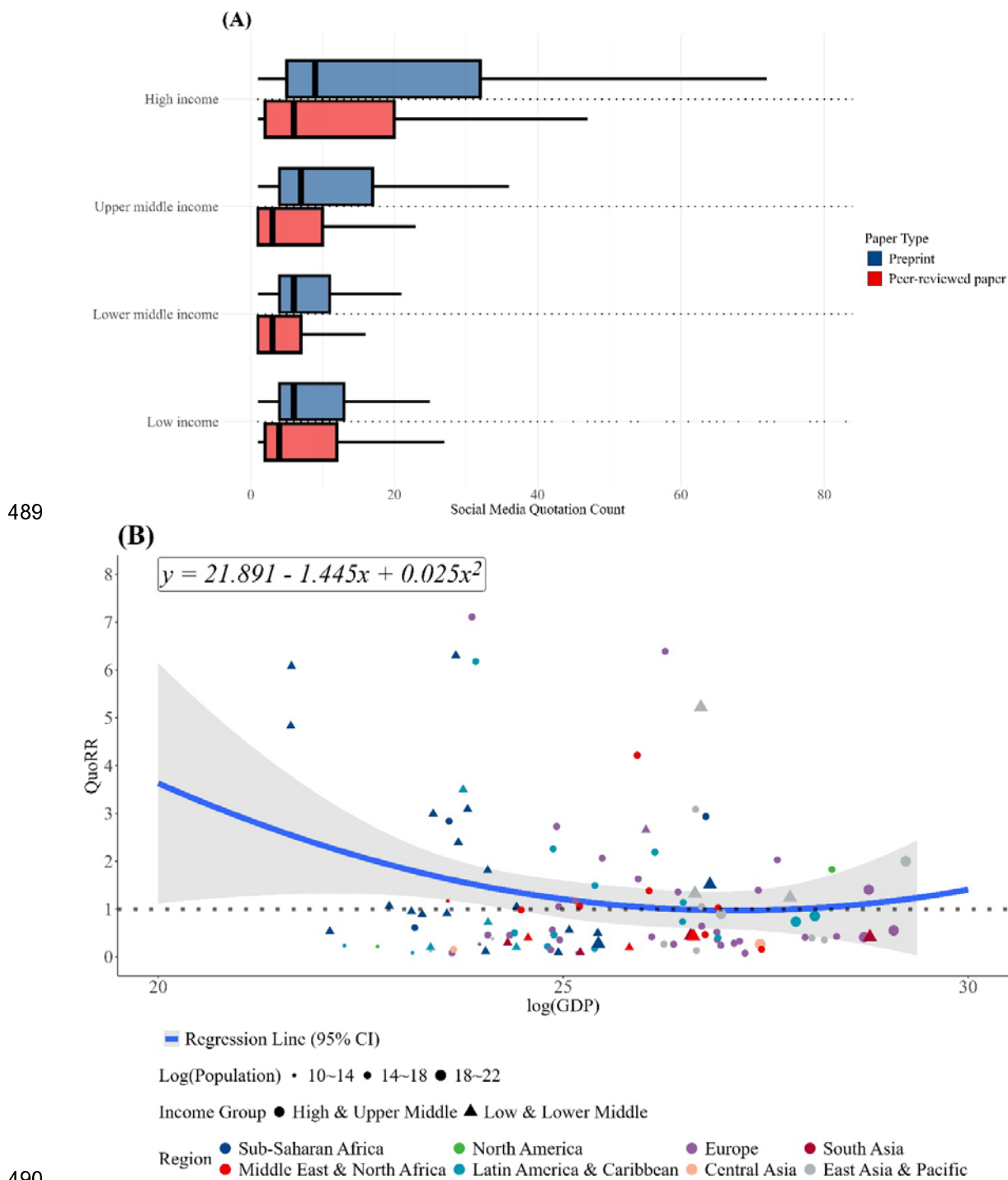
476 Figure 2 was made by first examining the count of papers published in each country based on the 1st author's  
 477 affiliation then calculating PubRR. Top 10% and bottom 10% outliers of PubRR values were excluded. In  
 478 Figure (a), X-axis and Y-axis of the scatter plot shows log(GDP) and PubRR value of each country. It also  
 479 shows Log(Population) value, Income Type (HMIC/LMIC) and Region of each country. The second-order  
 480 polynomial trend line is shown to be  $y = 30.38 - 1.99x + 0.03x^2$  ( $R^2 = 0.159$ ), and the Pearson correlation  
 481 coefficient  $-0.28$  with a  $p$ -value  $< 0.001$ . Figure (b) shows PubRR value on a geographical heatmap with  
 482 countries with missing PubRR values colored as gray.

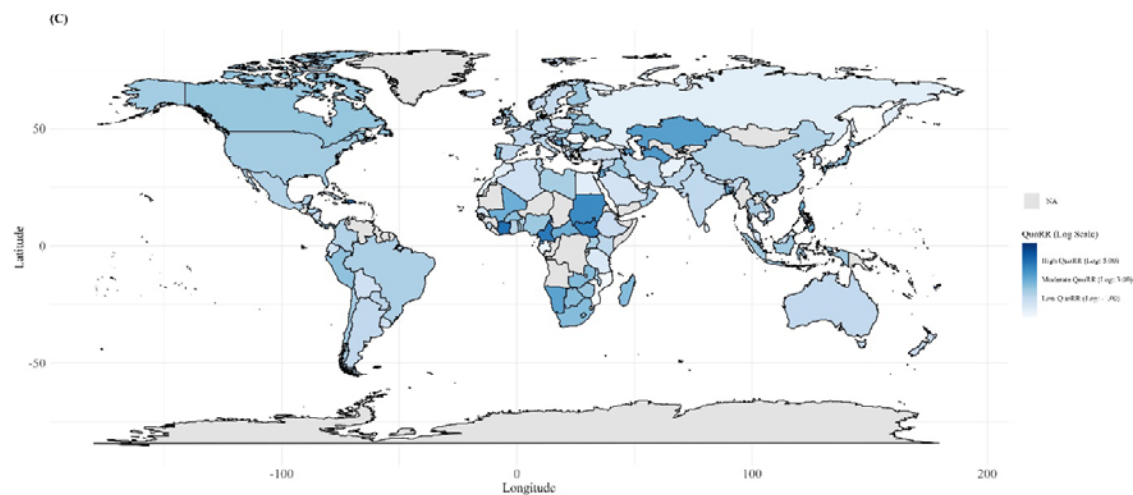
483 **Abbreviations:** GDP, Gross Domestic Product; PubRR, Publication Relative Ratio; CI, Confidence Interval;  
 484 Log, Common Logarithm; HMIC, High-Middle Income Countries; LMIC, Low-Middle Income Countries; NA,  
 485 Not Available.

486



487 **Figure 3. (A) Boxplot of Social Media Quotation Count by Income and Paper Type, (B) Scatter plot of the**  
 488 **country's GDP and QuoRR (C) Geographical Heatmap of QuoRR**





491

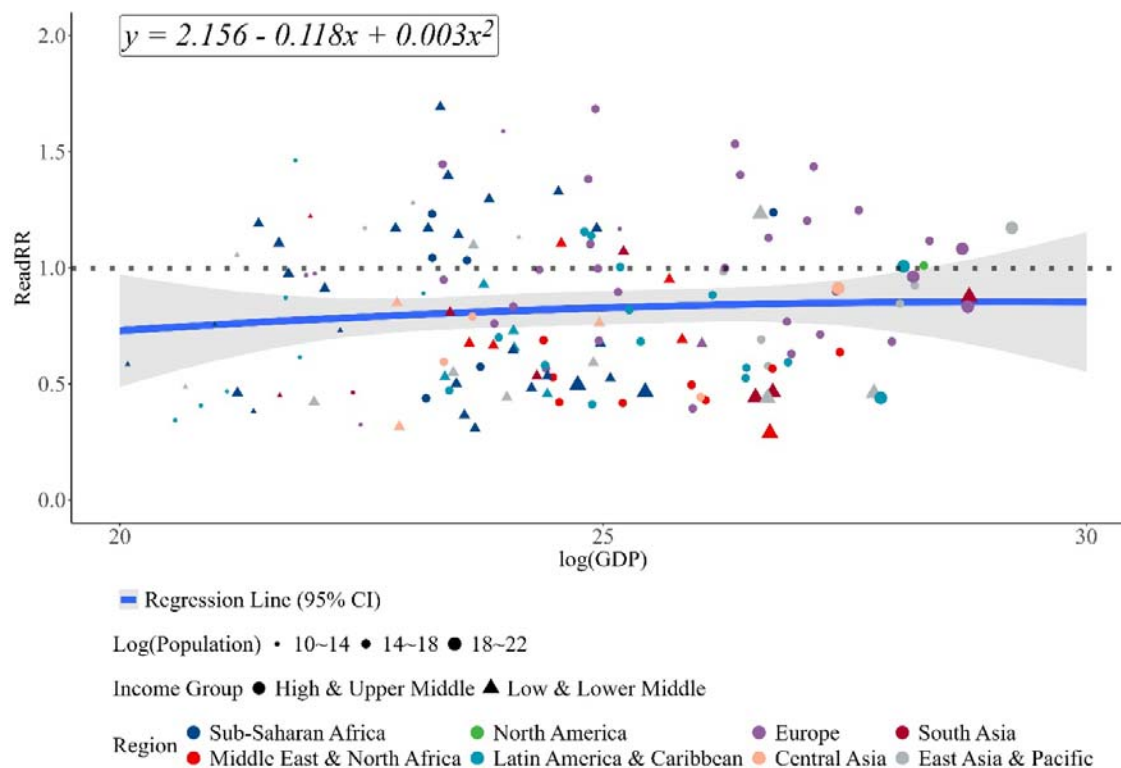
492 Figure 3 (a) is a set of boxplots for each paper type (preprints and peer-reviewed papers) based on the social  
493 media quotation counts of each paper. They are sub-grouped by the income group of the 1st author's country.  
494 Outliers were excluded when drawing these boxplots. Five-number summaries for each boxplot starting from  
495 the top are (1, 5, 9, 32, 72), (1, 2, 6, 20, 47), (1, 4, 7, 17, 36), (1, 1, 3, 10, 23), (1, 4, 6, 11, 21), (1, 1, 3, 7, 16), (1,  
496 4, 6, 13, 25) and (1, 2, 4, 12, 27). Figure 3 (b), (c) was made by examining the social media quotation counts of  
497 papers published in each country then calculating QuoRR. They are drawn in the same way that Figure 2 was  
498 drawn. The second-order polynomial trend line is shown to be  $y = 21.89 - 1.45x + 0.03x^2$  ( $R^2 = 0.03$ ), and  
499 the Pearson correlation coefficient  $-0.17$  ( $p$ -value = 0.046).

500 **Abbreviations:** GDP, Gross Domestic Product; QuoRR, Social Media Quotation Relative Ratio; CI, Confidence  
501 Interval; Log, Common Logarithm; HMIC, High-Middle Income Countries; LMIC, Low-Middle Income  
502 Countries; NA = Not Available.

503

504

505 **Figure 4. Scatter plot of the country's GDP and ReadRR**



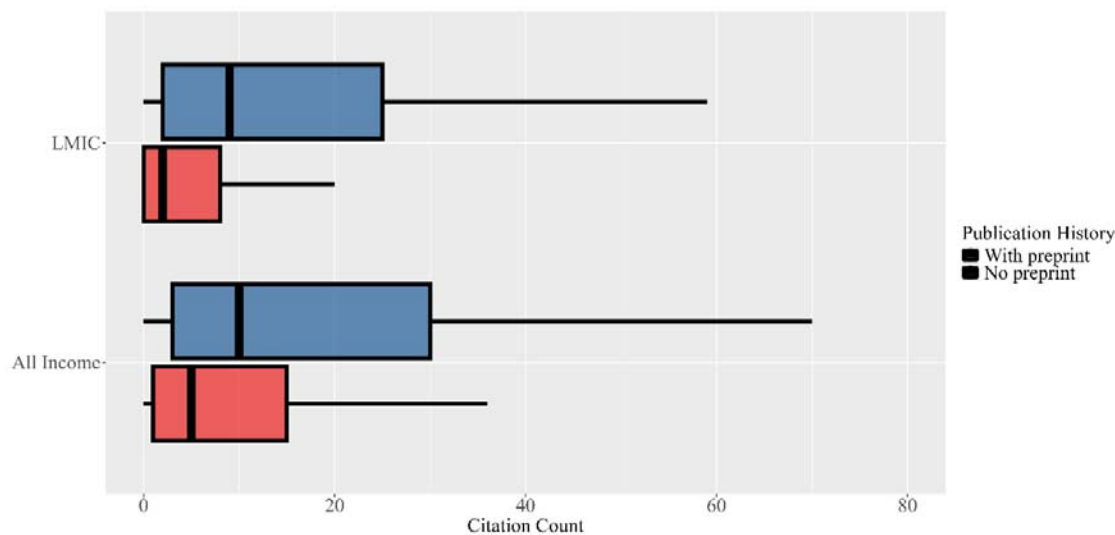
506

507 Figure 4 was made by first examining social media quotation counts of peer-reviewed papers and preprints left  
508 by X users from each country then calculating ReadRR. It is drawn in the same way that Figure 2 was drawn.  
509 The second-order polynomial trend line is shown to be  $y = 2.16 - 0.12x + 0.003x^2$  ( $R^2 = 0.01$ ), and the  
510 Pearson correlation coefficient -0.11 but with a non-significant p-value of 0.136.

511 **Abbreviations:** GDP, Gross Domestic Product; ReadRR, Social Media Readership Relative Ratio; CI,  
512 Confidence Interval; Log, Common Logarithm; HMIC, High-Middle Income Countries; LMIC, Low-Middle  
513 Income Countries.

514

515 **Figure 5. Boxplot of citation count by preprint publication history**



516

517 The citation count distribution of peer-reviewed papers with and without preprint publication history is shown.  
518 Outliers were excluded. Peer-reviewed papers with preprint history exhibit higher median and quartile citation  
519 counts. The Wilcoxon rank-sum test indicates a significant increase in citations for papers with preprint history  
520 (median: 10, Q1-Q3: 3-30) versus those without (median: 5, Q1-Q3: 1-15), with a p-value < 0.001. This trend is  
521 also confirmed in LMIC countries, with median citations of 9 (Q1-Q3: 2-25) for papers with preprint history  
522 compared to 2 (Q1-Q3: 0-8) for those without, supported by a p-value < 0.001.

523 **Abbreviations:** LMIC, Low-Middle Income Countries.

524

525 **Table 1. Coefficient of variation in inter-regional co-authorship network**

	<b>Mean</b>	<b>SD</b>	<b>CV</b>
<b>Preprint</b>	6833.6	13,300.3	1.95
<b>Peer-reviewed paper</b>	196,572.3	530,283.8	2.70

526 CV was calculated based on inter-regional co-authorships between all possible pairs of regions, involving 36  
527 unique combinations (8 regions taken 2 at a time), to reflect the weight of edges across both preprints and peer-  
528 reviewed papers. Our findings revealed a skewed distribution of international co-authorship in peer-reviewed  
529 papers, heavily favoring European countries. In contrast, preprints demonstrated a more balanced co-authorship  
530 pattern, as indicated by a CV of 1.95 for preprints compared to 2.70 for peer-reviewed papers, highlighting the  
531 greater concentration of collaborations in the latter.

532 **Abbreviations:** CV, Coefficient of Variation; SD, Standard Deviation.

533

534 **Table 2. Eigenvector centrality in inter-regional co-authorship network**

	Peer-reviewed paper	Preprint	CEI
Europe	1.000	1.000	1.00
North America	0.106	0.812	7.66
East Asia & Pacific	0.077	0.135	1.76
Latin America & Caribbean	0.026	0.103	3.90
Middle East & North Africa	0.040	0.088	2.19
Sub-Saharan Africa	0.012	0.055	4.53
South Asia	0.031	0.033	1.07
Central Asia	0.008	0.003	4.29

535 EVC distribution was computed using the nodes and edges of the network represented as an adjacency matrix in  
536 Figure 1. The CEI was calculated by dividing the EVC of preprints by that of peer-reviewed publications. This  
537 metric serves as a tool for direct comparison of collaboration intensity; They reflect the influence of each region  
538 on international collaboration by weighting collaborations with scientifically significant regions such as Europe,  
539 North America, and East Asia and Pacific. The regions with the largest increases in collaboration intensity for  
540 preprints when compared to peer-reviewed publications are North America (7.66), Sub-Saharan Africa (4.53),  
541 Central Asia (4.29), Latin America & the Caribbean (3.9), and the Middle East & North Africa (2.19).  
542 **Abbreviations:** EVC, Eigenvector Centrality; CEI, Co-authorship Enhancement Index.

543