1

2

3

4

Unreviewed science in the news:
The evolution of preprint media coverage from 2014-2021

Alice Fleerackers[1*] (https://orcid.org/0000-0002-7182-4061), Kenneth Shores[2] (https://orcid.org/0000-0003-3039-420X), Natascha Chtena[3] (https://orcid.org/0009-0006-3586-5372), Juan Pablo Alperin[3*] (https://orcid.org/0000-0002-9344-7439)

[1] Interdisciplinary Studies, Simon Fraser University, Vancouver, BC, Canada, [2] School of Education, University of Delaware, Newark, DE, USA, [3] School of Publishing, Simon Fraser University, Vancouver, BC, Canada

*Correspondence should be addressed to Alice Fleerackers and Juan Pablo Alperin, School of Publishing, Simon Fraser University, 515 West Hastings St. Vancouver, BC V6B 5K3, emails: hello@alicefleerackers.com and juan@alperin.ca

Author contributions (CREDIT):

A.F.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Writing - original draft, and Writing - review & editing.
J.P.A.: Conceptualization, Data curation, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Supervision, Writing - original draft, and Writing - review & editing.
K.S.: Formal analysis, Methodology, Software, Visualization, Writing - original draft, and Writing -

1

Peer Review: https://www.webofscience.com/api/gateway/wos/peer-review/10.1162/qss_a_00282

34   review & editing.

35   **N.C.:** Writing - original draft and Writing - review & editing.

Abstract

It has been argued that preprint coverage during the COVID-19 pandemic constituted a paradigm shift in journalism norms and practices. This study examines whether and in what ways this is the case using a sample of 11,538 preprints posted on four preprint servers—bioRxiv, medRxiv, arXiv, and SSRN—that received coverage in 94 English-language media outlets between 2014–2021. We compared mentions of these preprints with mentions of a comparison sample of 397,446 peer reviewed research articles indexed in the Web of Science to identify changes in the share of media coverage that mentioned preprints before and during the pandemic. We found that preprint media coverage increased at a slow but steady rate pre-pandemic, then spiked dramatically. This increase applied only to COVID-19-related preprints, with minimal change in coverage of preprints on other topics. The rise in preprint coverage was most pronounced among health and medicine-focused media outlets, which barely covered preprints before the pandemic but mentioned more COVID-19 preprints than outlets focused on any other topic. These results suggest that the growth in coverage of preprints seen during the pandemic may imply only a temporary shift in journalistic norms, including a changing outlook on reporting preliminary, unvetted research.

Keywords: journalism, preprints, COVID-19, altmetric, science communication, news

53   1.0 Introduction

54   On January 10, 2020, the World Health Organization published its first set of guidelines for

55   preventing and controlling a suspected "novel coronavirus (nCoV)" (WHO, 2020). Soon journalists

56   found themselves plunged into an unexpected crisis, with an out-of-control, little understood infectious

57   disease, and an influx of new scientific information to sift through and report on. Without much peer

58   reviewed literature to go on—especially in the early stages of the pandemic—many turned to preprint

59   servers to share urgent new information with the public (Fraser et al., 2021). The ensuing media

60   coverage of preprints seen during the pandemic has since been described as a complete rupture from past

61   reporting practices (e.g., Burke, 2021; Makri, 2021). Yet, empirical evidence supporting this assertion is

62   lacking. As noted in previous research, there is currently an absence of longitudinal investigations that

63   examine preprint coverage over time *and* which assess the impact of COVID-19 on journalistic practices

64   and norms (Fleerackers et al., 2023; van Schalkwyk & Dudek, 2022b). This study fills this gap by

65   examining how media coverage of preprints has evolved, both qualitatively and quantitatively, in the

66   lead up to, and during the first year of, the COVID-19 pandemic. Using Altmetric data, it examines

67   changes in the volume and nature of media coverage of 11,538 preprints posted between 2013 and 2021

68   on bioRxiv, medRxiv, arXiv, and SSRN—four of the most actively used servers used to share COVID-

69   19-related research (Waltman et al., 2021).

70   2.0 Literature review and research questions

71   2.1 Preprint media coverage before and during the COVID-19 pandemic

72   Preprints have been used extensively in physics, math, and computational science since arXiv

73   launched in 1991. However, scientists in the biological and medical fields have been more reluctant to

4

74   do so—that is, until recently (Puebla et al., 2021). The early months of the pandemic saw a sharp

75   increase in the volume of available COVID-19-related preprints (Funk, 2023; Horbach, 2020), with

76   preprint servers such as medRxiv and bioRxiv becoming key disseminators of pandemic research (Else,

77   2020; Vergoulis et al., 2021). One study (Kousha & Thelwall, 2020) found that preprints posted to arXiv

78   bioRxiv, medRxiv, and SSRN comprised 13.26% of the COVID-19 literature during March–April 2020,

79   while an analysis by Fraser et al. (2021) found that preprints posted to 16 servers (including the four

80   examined in this study) comprised almost 25% of the COVID-19-related research available from

81   January–October 2020. Studies have predicted that the use of pandemic-related preprints continued to

82   grow at a relatively stable rate throughout 2021 and 2022, although more research is needed to confirm

83   these predictions (Nane et al., 2023).

84   COVID-19-related preprints also gained traction within news media, receiving coverage in

85   diverse media outlets around the world (Fleerackers et al., 2021; Massarani et al., 2021; Massarani &

86   Neves, 2021; Simons & Schniedermann, 2023; van Schalkwyk & Dudek, 2022a). One study found that

87   more than a quarter of COVID-19-related bioRxiv and medRxiv preprints were mentioned in at least

88   one media story during the pandemic, while only about 1% of those on other topics received media

89   coverage (Fraser et al., 2021). Some journalists reported adopting novel practices to report on these

90   unreviewed studies, something they said they had never done before (Fleerackers et al., 2022a;

91   Massarani et al., 2021).

92   This media coverage of preprints seen during the COVID-19 pandemic has been described by

93   some journalists as a "paradigm shift" (Fleerackers et al., 2022a). Yet, while studies conducted during

94   the COVID-19 pandemic provide important evidence into how journalists covered preprints during the

95   evolving health crisis, little is known about whether journalists have covered preprints on other topics or

96   during other communication contexts. For example, Fraser et al. (2021)'s widely cited study is often

97   described as providing evidence that "During the pandemic, journalists…paid increased attention to

5

98   preprints" (Kwon, 2021), but the authors did not compare pandemic preprint coverage to pre-pandemic

99   levels. Instead, they provided evidence that COVID-19-related preprints received an outsized amount of

100   media attention, *relative to those on other topics posted to bioRxiv and medRxiv during the same time*

101   *period*—but not relative to preprints posted during different time periods or on different servers (Fraser

102   et al., 2021). One recent study begins to fill this gap through an examination of coverage of preprints by

103   seven German newspapers from 2018-2021 (Simons & Schniedermann, 2023). The authors identified

104   low and stable rates of coverage leading up to the pandemic, followed by a major surge in 2020 and

105   2021 that was driven by COVID-19-related preprints. However, it is unclear whether this trend is

106   reflective of other media outlets (e.g., those outside of Germany) and whether there are disciplinary

107   differences in coverage trends.

108   More broadly, although preprints made up a significant proportion of the COVID-19-related

109   literature available within the first months of the pandemic, it is unclear how media coverage of

110   preprints compares to coverage of peer reviewed research. One article found that the five COVID-19-

111   related research articles that received the most media coverage were all peer reviewed publications;

112   however, the analysis was descriptive and did not compare the volume of preprint coverage to that of

113   peer reviewed papers (Kousha & Thelwall, 2020). Another small study found no significant difference

114   in the amount of media coverage received by medRxiv preprints and peer reviewed publications about

115   COVID-19-related therapies that were posted between February 1–May 10, 2020 (Jung et al., 2021). A

116   study of South African media found that only 3% of stories mentioning COVID-19 research included a

117   mention of a preprint (van Schalkwyk & Dudek, 2022a). Besançon et al. (2021) used Altmetric to

118   examine news coverage of COVID-19-related preprints posted to arXiv, medRxiv, and bioRxiv between

119   January–July 2020, finding that these preprints received more coverage than the non-COVID-19-related

120   preprints posted to arXiv during the same time period. Again, coverage of preprints before the pandemic

121   period was not considered. Fraser et al. (2020) found that bioRxiv preprints submitted between

6

122 November 2013 and December 2017 received far less media coverage than either their peer reviewed

123 versions or a control set of peer reviewed articles that were never deposited to bioRxiv. Finally,

124 Waltman et al. (2021) found that, although some COVID-19-related preprints were highly reported on,

125 overall, news coverage of peer reviewed literature outstripped coverage of preprints. Unfortunately,

126 Waltman et al. (2021) did not report the average attention received per preprint vs peer reviewed article.

127 However, the authors did examine news coverage received by a sample of high-profile preprints and

128 their corresponding peer reviewed articles. For 45% of these preprint-article pairs, the preprint received

129 more than 20% of the total news attention; for 11% of the pairs, preprints received more than 80% of the

130 coverage (Waltman et al., 2021). Again, the authors did not compare these findings to rates of coverage

131 before the pandemic.

132 Collectively, these results provide some of the first evidence that preprints have historically

133 received less media coverage than peer reviewed research and that this trend may have started to shift

134 during the pandemic. However, given the mixed and incomplete body of evidence, several questions

135 remain unanswered. In particular, it is unclear whether the volume of preprint media coverage increased,

136 decreased, or remained relatively stable in the years leading up the pandemic—information that could

137 help shed light on whether preprint-based media coverage is likely to continue post-COVID-19. It is also

138 unclear whether any changes in coverage seen during the pandemic apply only to COVID-19-related

139 preprints or reflect a change in journalists' willingness to use preprints in general. As such, to examine

140 whether the pandemic has truly introduced a "paradigm shift" in journalistic practice, this study uses a

141 sample of preprints that received coverage in English-language media between 2014–2021 to examine

142 the following research questions:

143 *RQ1: Has the share of preprint coverage in the media increased during the COVID-19*

144 *pandemic?*

7

145   *RQ2: Do changes in media coverage of COVID-19-related preprints extend to coverage of*

146   *preprints on other topics?*

147   2.2 Preprint media coverage in an evolving media landscape

148   It is also unclear from previous research which types of media outlets have driven media

149 coverage of preprints and whether this has changed as a result of the pandemic. Journalism has evolved

150 in important ways in the years leading up to the COVID-19 crisis, with financial pressures, shrinking

151 news audiences, and changes to the digital communication landscape contributing to declines in

152 specialized science journalism around the world (Saari et al., 1998; Schäfer, 2017). These declines have

153 likely influenced the amount of media coverage that research articles—including preprints—receive, as

154 outlets specializing in science appear to cover more research than general interest publications (Wihbey,

155 2017). In addition, an array of actors who have historically been considered "peripheral," or outside of

156 journalism, have entered the field, including bloggers, news aggregators, and other alternative outlets

157 (Hermida, 2019; Schapals, 2022; Stocking, 2019). These peripheral actors may not always adhere to the

158 established norms and practices that shape media coverage at traditional—or "legacy"—outlets (e.g.,

159 Harrison et al., 2020; Hurley & Tewksbury, 2012), which may affect how or whether they cover

160 preprints. For example, journalists working at peripheral outlets may not be expected to adhere to

161 professional journalism resources, such as the AP Style Guide, which recommend avoiding research that

162 has not been peer reviewed (Froke et al., 2020; Haelle, 2020). Yet, both peripheral and legacy outlets

163 actively covered COVID-19-related preprints during the early months of the pandemic (Fleerackers et

164 al., 2021). Similarly, outlets which publish content but are not considered journalism, such as university

165 websites and press release distribution services, may also contribute to mobilizing preprint research. For

166 example, the Science Media Centre in Germany—a non-journalistic outlet that provides science

167 journalists with access to research and expert perspectives—began sharing roundups of newly posted

8

168 preprints during the pandemic (Broer, 2020; Broer & Pröschel, 2022). Again, however, any evidence

169 about the nature of non-journalistic outlets reporting on preprints is limited to the pandemic period. As

170 such, our third research question asks:

171 *RQ3: Have changes in media coverage of preprints occurred similarly across media outlets?*

172 3.0 Method and materials

173 To identify media coverage, this study relies on data from Altmetric,[1] a company that tracks

174 mentions of research outputs across a range of digital media, including news media. Research suggests

175 that Altmetric's "Mainstream Media" category is a relatively reliable source of data but only when

176 working with a predefined list of English-language media outlets (Fleerackers et al., 2022a; Ortega,

177 2020b, 2020a). In addition, because Altmetric regularly updates both the list of media outlets[2] and

178 research outputs[3] it tracks, the volume of media coverage it collects may vary over time in ways that are

179 unrelated to actual changes in news reporting. For these reasons, we decided to gather two datasets:

180   1. A primary dataset comprising news mentions of bioRxiv, medRxiv, arXiv, and SSRN preprints;

181   2. A comparison dataset comprising news mentions of peer reviewed research indexed in the Web

182   of Science (WoS).

183 3.1 Identifying and characterizing media outlets that frequently cover research

184 Data were queried from local snapshots of the Web of Science and Altmetric Databases housed

185 at the Observatoire des sciences et des technologies (OST) on January 30, 2023.[4] Data filtering and

---

[1] altmetric.com

[2] https://help.altmetric.com/support/solutions/articles/6000235999-news-and-mainstream-media

[3] https://www.altmetric.com/about-our-data/how-it-works-2/

[4] https://www.ost.uqam.ca/

9

186   cleaning were performed using the Python pandas package (The Pandas Development Team, 2023). To

187   identify our predefined set of media outlets, we queried a snapshot of the Altmetric database from June

188   3, 2021, for news mentions of all WoS research outputs associated with a digital object identifier (DOI).

189   We restricted our search to mentions of research outputs that had been published in 2013 or later and

190   that were mentioned in news stories between January 1, 2014, and June 3, 2021. We then filtered for

191   outlets that consistently covered a high volume of research, defined for the purposes of this study as

192   outlets that mentioned at least 100 WoS research items per year from 2014–2020. We manually checked

193   the resulting 128 media outlets by visiting the URLs for their home pages provided by Altmetric. After

194   excluding 25 outlets that were not written in English, five that were not tracked by Altmetric from 2021–

195   2022 (e.g., because they had changed their domain names), three whose URLs did not resolve, and one

196   with all misidentified mentions, we were left with a final sample of 94 outlets.

197   Next, we applied a coding protocol adapted from Hermida & Young (2019) to characterize the

198   nature of these media outlets. We analyzed each outlet's main topical focus (e.g., science and

199   technology, health and medicine, general news, etc.) and assessed whether it was best described as

200   legacy journalism (i.e., staffed by professional journalists who adhere to traditional journalistic norms),

201   peripheral journalism (i.e., staffed by individuals who have traditionally worked outside of journalism

202   and who adhere to emerging or alternative norms), or non-journalism (i.e., organizations such as

203   universities, press release services, or academic journals that do not produce journalism). A detailed

204   version of the coding protocol, including examples, is available from Fleerackers and Fagan (2022).

205   Coding was performed by researchers with professional journalism experience: the lead author and a

206   research assistant who was not aware of the study objectives (cf. Hermida & Young, 2019). The two

207   coders independently explored the media outlets' websites, examining their content, Mission Statement,

208   and, if available, other relevant pages (e.g., Masthead, Editorial Guidelines, Code of Conduct). The

209   coders compared their coding and resolved any discrepancies through discussion, and, if needed, by

10

210   consulting an outside researcher (also a former journalist). Such double coding approaches are

211   appropriate when data are not very numerous (Krippendorff, 2004), as in the present study. Results of

212   the final coding are reported in aggregate in Table 1; coding for the full list of outlets is available at

213   (Alperin, Fleerackers, et al., 2023).

214   **Table 1.** Nature of media outlets that frequently cover Web of Science research

Topic	N	Type	N
General News	38	Legacy	49
Science/Technology	28	Peripheral	35
Health/Medicine	15	Non-journalism	10
Other	13		
Total	94		94

215   3.2 Gathering news mentions of preprint research

216   We gathered news mentions of preprints from four servers—bioRxiv, medRxiv, arXiv, and

217   SSRN—because these servers were highly used for sharing COVID-19-related preprints (Waltman et

218   al., 2021). These servers were also launched at different times (bioRxiv in 2013, medRxiv in 2019,

219   arXiv in 1991, and SSRN in 1994), with different disciplinary scopes, and have seen different levels of

220   uptake among scholars (Puebla et al., 2022), providing us with a diverse sample of preprints for our

221   analysis. We queried Altmetric for mentions of preprints from these servers in stories published by the

222   94 outlets since January 1, 2014. This yielded a total of 40,039 mentions of 15,041 preprints across

223   31,258 news stories. For each of these preprints, we gathered the publication dates from the arXiv and

224   Crossref APIs using the Python arxiv and habanero packages (Chamberlain, 2020; Schwab, 2021).

225   Next, because previous research suggests publication date metadata can often be incorrect or

226   incomplete (Haustein et al., 2015), we manually checked subsamples of our data and compared the

11

227 publication dates provided by Crossref, the arXiv API, and Altmetric. The most reliable publication date

228 for each server was retained for analysis. For bioRxiv and medRxiv, this was the DOI creation date (i.e.,

229 the date that the DOI for the preprint was deposited in Crossref); for arXiv, it was the date provided by

230 the arXiv API; and for SSRN, it was either the "first posted on" date provided by Altmetric or Crossref's

231 DOI creation date, whichever came first. We removed 3,619 preprints that were published before 2013,

232 as these publication dates were particularly unreliable (perhaps because Altmetric started tracking

233 mentions partway through 2012 and thus has incomplete data for previously published outputs)[5]. Even

234 after excluding these preprints and selecting the most reliable publication date for each server, we noted

235 that publication dates for *arXiv* and *SSRN* sometimes differed from the dates visible on the server web

236 page by a few days—a limitation that we kept in mind during data cleaning and analysis.

237 We made several further exclusions to ensure that the mentions in our dataset were mentions of

238 true preprints (i.e., rather than postprints or journal versions of preprints). First, we removed 165

239 mentions of postprints, which we defined as preprints that were posted on the same day, or after, their

240 journal versions were published. Because, as mentioned above, publication dates for preprints were

241 often incorrect by a few days, we excluded an additional 332 mentions of preprints with a publication

242 date within seven days of the journal version's publication date (i.e., suspected postprints). We also

243 removed 327 mentions of preprints in news stories that were published before the preprint was first

244 posted, using a five-day cut off to allow for the slight inconsistencies we identified in the publication

245 metadata. Because Altmetric does not disambiguate between preprints and journal versions for some

246 preprint servers[6,7] and may thus erroneously include some mentions of peer reviewed research, we

247 removed 3,547 mentions in news stories published after the peer reviewed journal version of the preprint

---

[5] NISO Altmetrics Working Group C "Data Quality" – Code of Conduct Self-Reporting Table
[6] https://help.altmetric.com/support/solutions/articles/6000240580-merging-preprints-and-final-published-versions
[7] https://www.altmetric.com/about-our-data/how-it-works-2/

12

248 was published, again using a five-day margin. While this approach may have removed some true

249 mentions of preprints, these false removals are likely limited, as journalists strive to ensure their stories

250 are timely and relevant (Rosen et al., 2016; Shoemaker & Reese, 1996) and seldom cover research

251 outputs more than a few weeks after initial publication (Maggio et al., 2017). Finally, we removed an

252 additional 1,021 duplicate news mentions (where the same preprint was mentioned in the same story

253 more than once). In total, filtering led to the exclusion of 9,081 mentions (22.5% of the original dataset).

254 The code used for filtering has been made publicly available (Alperin, Shores, et al., 2023). Our final

255 preprint sample comprised 31,028 mentions of 11,538 preprints by the 94 outlets in our sample (Alperin,

256 Fleerackers, et al., 2023).

257 ## 3.3 Gathering news mentions of peer-reviewed research

258 We downloaded all the mentions of WoS research from our 94 outlets (i.e., those described in Section

259 3.1), resulting in 1,657,202 mentions of 466,138 distinct research outputs. From these, we filtered

260 156,187 mentions of research articles that were published prior to 2013, 579 mentions that were already

261 included in the preprint data, and 14,482 duplicate mentions (where an article was mentioned in the

262 same news story more than once). In total, filtering led to the exclusion of 170,669 mentions (10.3% of

263 original dataset).

264 The final journal research sample comprised 1,486,533 mentions of 397,446 distinct peer

265 reviewed research outputs by the 94 outlets (Alperin, Fleerackers, et al., 2023).

266 ## 3.4 Identifying news mentions of COVID-19 research

267 To identify COVID-19-related preprints and WoS outputs, we searched for the presence of the

268 following COVID-19-related keywords in the outputs' titles using R version 4.3.0 (2023): *coronavirus,*

269 *covid-19, sars-cov, sars-cov-2, ncov-2019, 2019-ncov, hcov-19, sars-2, pandemic, covid, Severe Acute*

13

270 *Respiratory Syndrome Coronavirus 2, 2019 ncov*. These keywords were a combination of those used by

271 Fraser et al. (2021) and those listed in the National Library of Medicine's search strategy for identifying

272 COVID-19-related literature (Chen et al., 2020). We also added the term "pandemic," which wasn't

273 included in either of these lists of keywords but is likely used in many COVID-19 titles. As some

274 keywords (e.g., "pandemic") may have been used in non-COVID-19 contexts, we also filtered for

275 research published in 2020 or later when identifying COVID-19-related research.

276 3.5 Statistical analyses

277   Statistical analysis was performed using Stata version 17 (StataCorp, 2021). The Stata script

278 used for the following analysis has been made publicly available (Alperin, Shores, et al., 2023).

279 Throughout our analyses, we examined changes in preprint media coverage in terms of proportions,

280 rather than counts. Specifically, we compared mentions of preprints against mentions of all research in

281 our sample (i.e., mentions of preprints *and* WoS research). Doing so allowed us to control for any

282 fluctuations in the volume of preprint mentions that were created by changes in Altmetric's approach to

283 identifying research mentions during the study period, rather than the result of changing journalistic

284 practices. For ease of reading, we use the term "share of preprint mentions" to refer to the proportion of

285 all research mentions that focused on preprints and "share of WoS mentions" to refer to the proportion

286 that focused on WoS research.

287   To answer RQ1, we created a model (Equation (1)) to estimate the degree to which medRxiv and

288 COVID-19 contributed to changes in the volume of media coverage of preprints after 2019.

289 Disentangling any change in preprint coverage due to the launch of the server and the onset of the

290 pandemic was necessary as the creation of medRxiv preprints in 2019 (Kaiser, 2019) coincided closely

291 with the start of the COVID-19 era. As such, in Equation (1), we estimated an ordinary least squares

292 (OLS) regression of a binary indicator (Y_{it}) coded as 1 if the news mention (*i*) referenced a preprint and

14

293 coded as 0 otherwise against time (t), encoded as linear days since Jan 1, 2014 and allowed to be

294 identified with 3rd-order polynomial trends (β_1 through β_3), with each vector of 3rd-order polynomial

295 terms estimated in both the pre-COVID-19 era and COVID-19 era (α_0 interacted with the vector of time

296 trends). We differentiated pre-COVID-19 from COVID-19 era mentions through a binary indicator,

297 coded as 1 if the preprint was mentioned in a news story published after January 10, 2020 (i.e., when the

298 WHO first used the term "2019-nCoV" to describe the novel coronavirus; WHO, 2020), and coded as 0

299 otherwise. We modeled the period between the first news mention of a medRxiv preprint (i.e., on July

300 23, 2019, which postdates the launch of the site on June 25, 2019 by about one month) and the WHO's

301 statement as a linear intercept shift (β_4). In practice, this variable allowed us to differentiate the change

302 in preprint mentions that occurred with the introduction of medRxiv before (but close to the onset of)

303 COVID-19 from the effect of COVID-19 itself. Similarly, we modeled the mentions of preprints with

304 titles that included COVID-19-related language (i.e., "sars-cov-2" or a related term) as a linear intercept

305 shift (β_5). This last variable is important, as it allowed us to differentiate the change in preprint

306 mentions for COVID-19-related topics in the media from changes in preprint prints in the COVID-19-

307 era but not about COVID-19 topics. Lastly, to adjust for seasonality and periodicity effects we

308 controlled for week-of-year intercepts (γ_{wy}; e.g., first week of 2014) and day-of-month effects (δ_{md};

309 e.g., Tuesdays in January). In practice, controlling for periodicity and seasonality had little effect on

310 model parameters but allowed us to rule out correlations between period effects and the onset of

311 COVID-19. The error term (ε_{it}) is robust to heteroskedasticity.

312 **Equation (1)**

313

314 $Y_{it} = \beta_0 + \alpha_0 CovidEra \times (\beta_1 Days + \beta_2 Days^2 + \beta_3 Days^3) + \beta_4 medRxiv + \beta_5 CovidLang + \gamma_{wy} + \delta_{md} + \varepsilon_{it}$

315

316 Next, we estimated separate OLS regressions that allow us to test whether changes in preprint

317 mentions vary across (a) preprint servers, (b) media outlets focused on different topics, and (c) media

15

318   outlets of different types. Because preprint servers necessarily represent preprint mentions, we discarded

319   mentions of articles from WoS and collapsed the data so that we could observe counts of preprint

320   mentions by day and identify any changes in these counts among the four servers (RQ2). To identify

321   changes among the four media outlet topics and three outlet types, respectively, we kept the data as

322   described previously, with each row representing a unique news mention of a preprint or WoS article.

323   To identify changes in the share of preprint mentions across the four media outlet topics (RQ3), we

324   focused on the three most prevalent topics in our sample—Health/Medicine, General News, and

325   Science/Technology—and an "Other" category that included a variety of other topics (e.g., Business,

326   Lifestyle, Explicit Point-of-View).

327   Because we are exploring heterogeneity across servers, topics, and types, we simplified the

328   regression Equation (1) by replacing the linear, quadratic, and cubic $Days$ variables with month-by-year

329   fixed effects (λ_{my} in Equation (2) below). These fixed effects control for time trends non-parametrically

330   in a similar way as in Equation (1) but without the need to directly identify the time effects (i.e., these

331   time effects are partialled from the regression equation as "nuisance parameters"). Our estimation

332   equation for these heterogeneous changes therefore appears as follows:

333   **Equation (2)**

334    $Y_{it} = \beta_0 + \beta_4 medRxiv + \Sigma^J_{j=1}\alpha_j \times \beta_5 CovidEra + \Sigma^J_{j=1}\alpha_j \times \beta_6 CovidLang + \lambda_{my} + \gamma_{wy} + \delta_{md} + \varepsilon_{it}$

335   where the key difference is that we identify changes in the pandemic era without and with

336   COVID-19-related titles ($\alpha_j \times \beta_5$ and $\alpha_j \times \beta_6$, respectively) for the four preprint servers (j=1 through

337   J=4), the four outlet topics (j=1 through J=4), and three outlet types (j=1 through J=3).

16

338 4.0 Results

339 4.1 Has the share of preprint coverage in the media increased during the COVID-19

340 pandemic?

341   Our models suggest that the annual number and share of preprint mentions increased slowly from

342 2014–2019, then increased dramatically in 2020-2021 (Table 2, Figure 1). However, even during the

343 pandemic period, preprint mentions made up only a small subset of media coverage of research, at less

344 than 5% of all mentions of research. We also saw evidence of a shift in which servers received the most

345 attention during the pandemic. Before the pandemic, most mentions of preprints cited preprints posted to

346 arXiv or SSRN; yet during the pandemic, bioRxiv and medRxiv became the most frequently mentioned

347 servers.
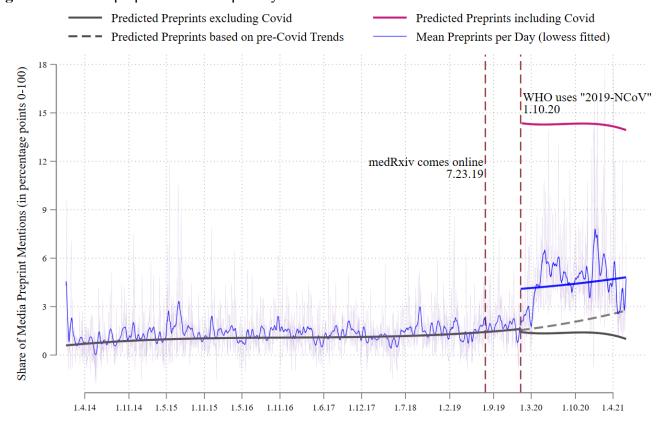
348 **Table 2.** Number and share of preprint mentions

349

Year	Number of WoS mentions	Number of preprint mentions	arXiv preprints	SSRN preprints	bioRxiv preprints	medRxiv preprints	Total preprint proportion
2014	98,580	753	0.34%	0.41%	0.01%	N/A	0.76%
2015	122,641	1,263	0.42%	0.56%	0.04%	N/A	1.02%
2016	192,414	2,115	0.44%	0.56%	0.09%	N/A	1.09%
2017	214,714	2,412	0.42%	0.60%	0.09%	N/A	1.11%
2018	218,946	2,597	0.48%	0.58%	0.12%	N/A	1.17%
2019	231,503	3,275	0.55%	0.67%	0.16%	0.01%	1.39%
2020	279,737	12,484	0.46%	0.56%	0.85%	2.40%	4.27%
2021*	127,998	6,129	0.50%	0.29%	1.27%	2.51%	4.57%
Total	1,486,533	31,028					

| % of all research mentions | 0.46% | 0.55% | 0.35% | 0.68% | 2.04% |
| % of all preprint mentions | 22.62% | 26.93% | 16.95% | 33.49% | 100% |

350   * partial year

351   **Figure 1.** Share of preprint mentions per day.

352
353   NB: Figure shows the average proportion of preprint mentions in the media per day beginning Jan 1, 2014, and
354   ending June 3, 2021. The fluctuating faded purple line plots residualized mean share of preprint mentions per day,
355   controlling for week-of-year effects (e.g., first week of 2014, second week of 2014, etc.) and day-by-month effects
356   (e.g., Tuesdays in January); the fluctuating blue line is the local linear regression (lowess) fitted line of those data.
357   The solid gray line is the predicted share of daily preprint mentions, with prediction based on a 3rd-order polynomial
358   function, controlling for week-of-year effects and day-of-month effects. We estimate an intercept shift for the period
359   before COVID-19 and after medRxiv was introduced on July 23, 2019, and an intercept shift for mentions of
360   preprints with titles that include COVID-19-related language (e.g., "novel coronavirus" or a related term). The solid
361   fuchsia line represents this last effect and is the estimated change in mean preprint mentions per day among all
362   COVID-19-related mentions of research (preprints and WoS publications). The dashed gray line is the predicted
363   share of daily preprint mentions based on trends observed in the period prior to the emergence of COVID-19. The
364   solid blue line describes the proportion of preprint mentions in the COVID-19 era and is the weighted average of
365   the fuchsia and gray lines. See Equation 1 for details.
366

18

367   With respect to medRxiv preprints, we found that the onset of COVID-19 increased the share of

368 preprint mentions in the media, beyond any increase due to the launch of the server in 2019.

369 Specifically, we estimate that, prior to the introduction of medRxiv and COVID-19, the share of

370 preprints mentioned in the media was increasing at a glacial pace (an *annual* rate of 0.21 percentage

371 points; p-value<0.000; 95% CI [0.13 - 0.29]; see solid gray line, Figure 1). When medRxiv was

372 introduced, the share of preprint mentions did not change (estimated decrease=0.005 percentage points;

373 p-value=0.957; 95% CI [-0.17 - 0.16]). In contrast, the share of preprint mentions increased by an

374 estimated 2.58 percentage points after the onset of the pandemic (p-value<0.000; 95% CI [2.45 - 2.70];

375 see solid blue line). This significant but modest increase applied to all preprint mentions, but masks

376 large differences in the proportion of preprint mentions between COVID-19-related and non-COVID-

377 19-related research during the pandemic.

378   Indeed, our model strongly suggests that preprints played a far greater role in media coverage of

379 COVID-19 specifically rather than in coverage of other topics. This can be seen from the "COVID-19"

380 line (in fuchsia) in Figure 1, which represents the estimated share of preprint mentions among all the

381 mentions of COVID-19-related research (i.e., both preprints and WoS articles that included COVID-19-

382 related language in the title). We estimated an increase in these COVID-19-related preprint mentions of

383 12.94 percentage points (p-value<0.000; 95% CI [12.84 - 13.04]), a large increase relative to predicted

384 preprint mentions based on pre-COVID-19 trends (gray dotted line). We explore coverage of non-

385 COVID-19 preprints in more detail in Section 4.2.

386   We further tested whether any changes in the share of preprint mentions seen during the

387 pandemic could be linked to changes in mentions of WoS research during this period. We implemented

388 this test by comparing growth rates of news mentions for preprints and WoS research over time. Given

389 that preprint mentions comprised only about 2 percent of all mentions in our sample and to place

390 preprint and WoS mentions on a common y-axis, we plotted preprint and WoS mentions as growth rates.
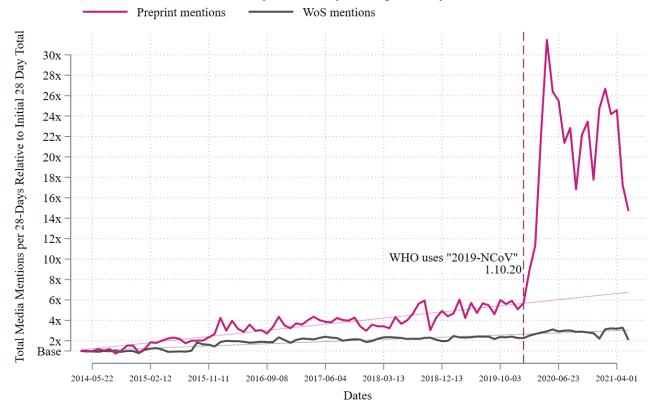
19

391   Growth rates for preprint and WoS mentions were each calculated using the total number of mentions in

392   the first 28 days of our data beginning with Sunday (i.e., January 5, 2014). These mentions in the first 28

393   days comprised our "base rate," and the total number of mentions in each sequential 28 days were then

394   scaled by that base rate.

395   Here, we find that the rise in the share of preprint mentions that took place during the pandemic

396   was not simply an artifact of a decrease in WoS mentions. As can be seen from Figure 2, WoS mentions

397   increased by about 2.3 percentage points between May 2014 to September 2019, and this pace of growth

398   remained relatively unchanged after COVID-19 began and started to garner media attention. In contrast,

399   preprint mentions had increased by about 5.7-fold by the time of the WHO's announcement about

400   "2019-nCoV" in January 2020, but skyrocketed to a 30-fold increase at the height of the pandemic in

401   May 2020. This figure thus shows that the increase in the proportion of preprint mentions during the

402   pandemic era was driven almost entirely by an increased number of preprint mentions and not a decrease

403   in the number of WoS mentions.

404   **Figure 2.** Growth rates of mentions for preprints and Web of Science (WoS) articles over time.

20

405
406   NB: Figure shows the number of preprint and Web of Science (WoS) mentions in the media per 28 days beginning
407   the first 28 days of March 2, 2014, and ending June 3, 2021. We select the first 28 days starting with March 2, 2014,
408   for stylistic reasons, as preprint mentions declined slightly between January 2014 and March 2014; however,
409   starting on January 1 does not change any of the findings we discuss below. Because preprint mentions only
410   comprise about two percent of all news mentions in our data and to put these counts on a common y-axis, we scale
411   monthly (28 day) preprint and WoS mentions by the number of mentions for preprints and WoS, respectively, by
412   the initial number of mentions in the first 28 days of our sample beginning March 2, 2014. Thus, subsequent monthly
413   mentions are relative to this base period. For example, a 28-day news mention count of "2x" means that news
414   mentions in that 28-day period were two times larger than news mention counts from March 2, 2014 to March 29,
415   2014 (i.e., the initial 28 day period). The fuchsia and gray lines indicate 28-day preprint and WoS news mentions,
416   respectively, in this relative metric.

417   4.2 Do changes in media coverage of COVID-19-related preprints extend to coverage of

418   preprints on other topics?

419   Our results suggest that the onset of the pandemic not only increased media attention to COVID-

420   19-related preprints but may have also decreased attention to preprints on other topics. Among all

421   research that excluded COVID-19-related language (solid gray line, Figure 1), we found that the share

422   of preprint mentions during the pandemic decreased by 0.18 percentage points, although this decrease

21

423   was not significant (p-value=0.129; 95% CI [-0.42 - 0.05]). Model-based estimates suggest that by June

424   3, 2021, if the pandemic had not occurred, we would have expected the share of preprint mentions to be

425   2.58 percentage points (dashed gray line, Figure 1); yet the observed share of non-COVID-19-related

426   preprint mentions comprised only 0.86 percentage points of all news mentions, a difference of 1.71

427   percentage points from what would have been expected (p-value<0.000; 95% CI [1.13 - 2.31]). This last

428   result suggests that the pandemic may have shifted media attention away from preprints about non-

429   COVID-19-related topics by modest amounts. In effect, our results suggest that COVID-19-related

430   preprint mentions eclipsed pre-pandemic preprint mentions.

431   Looking at the number of preprint mentions by server, we observed that there was no increase in

432   non-COVID-19-related preprint mentions in the pandemic for any server (Figure 3). All point estimates

433   were trivially small—about 0.7 to 1.8 fewer mentions per day, on average—and not statistically

434   significantly different from zero (p-values range from 0.217 to 0.699). For articles that included

435   COVID-19-related language in the titles, there was an average increase in daily news mentions of

436   bioRxiv and medRxiv preprints—of 6.2 (p-value<0.000; 95% CI [5.34 - 6.95]) and 19.2 (p-value<0.000;

437   95% CI [18.25- 20.08]), respectively—and a significant decrease in average daily news mentions for

438   arXiv and SSRN—of -2.7 (p-value<0.000; 95% CI [-3.75 – -1.68]) and -1.6 (p-value<0.000; 95% CI [-

439   2.69 – -0.51]) mentions per day, on average. In total, for the 511 days in the pandemic era in our sample,

440   this amounted to an increase of about 9,800 total mentions of medRxiv preprints and 3,170 total

441   mentions of bioRxiv preprints.

442   It is important to note that the declines in mentions of arXiv and SSRN preprints were only

443   significant for preprints that included a COVID-19-related keyword in the title. That is, the media were

444   less likely to mention preprints from these servers that were about COVID-19; instead, when

445   communicating about pandemic research, they tended to mention bioRxiv or medRxiv preprints. These

446   results could suggest that the media drew on the servers they expected would house the research most

22

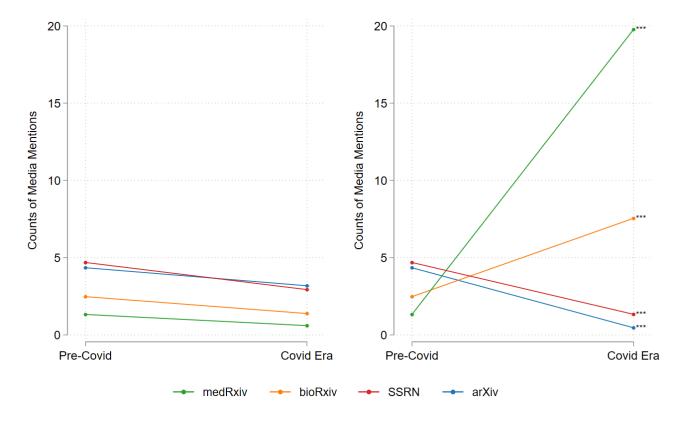447 relevant to their area of interest. It also suggests that COVID-19-related coverage tended to focus on

448 medical aspects of the pandemic and less so on social or economic aspects.

449 **Figure 3:** Change in the average daily count of preprint mentions, by server.

(a) Excluding COVID-19-related Language   (b) Including COVID-19-related Language

452 NB: Model parameters are from Equations (2). Statistical significance is determined based on the test of whether
453 pandemic era counts are different from pre-pandemic counts for mentions of preprints with titles that exclude
454 COVID-19-related language (panel a) and include COVID-19-related language (panel b). Key: * 10%
455 significance; ** 5% significance; *** 1% significance

456 4.3 Have changes in media coverage of preprints occurred similarly across media outlets?

457 Finally, we tested how preprint mentions changed across media outlets with four different topical

458 foci (i.e., General News, Science/Technology, Health/Medicine, Other) or of different types (i.e., legacy,

459 peripheral, or non-journalism). For mentions of COVID-19-related research, we found that outlets in all

460 four topic categories increased their preprint coverage dramatically during the pandemic, but to different

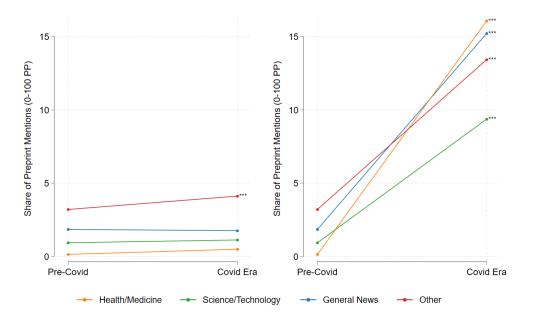461 extents. Increases ranged from 8.3 percentage points (Science/Technology) to 15.6 percentage points

23

462 (Health/Medicine) and were all statistically significant (p-value<0.000 for all coefficients) (Figure 4).

463 Changes in the share of mentions for non-COVID-19-related preprints were trivial, with only the

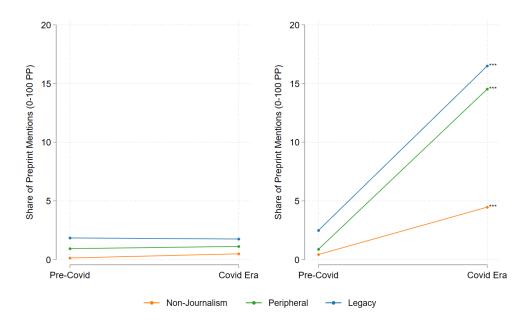464 "Other" category seeing a small but statistically significant increase (0.9 percentage points).

465 **Figure 4.** Change in the average daily share of preprint mentions, by topic and outlet.

4a. Media Outlet Topics

4b. Media Outlet Types

(a) Excluding COVID-19-related Language   (b) Including COVID-19-related Language

24

466 NB: Model parameters are from Equations (2). Statistical significance was determined by testing whether pre-pandemic share
467 of preprint news mentions was different from those during the pandemic, both among mentions of research with titles that
468 excluded COVID-19-related language (panel A) and those that included COVID-19-related language (panel B). Key: * 10%
469 significance; ** 5% significance; *** 1% significance
470
471   Similarly, none of the outlet types (i.e., legacy, peripheral, non-journalism) saw a statistically

472 significant increase in the share of preprint mentions of non-COVID-19-related articles, and the

473 coefficients themselves were trivially small, never reaching 1 percentage point. However, just as with

474 the topic-based data, all three outlet types increased the share of mentions of COVID-19-related

475 preprints after the WHO announcement in 2020 (estimates ranged from 4 to 14 percentage points for

476 non-journalism and legacy outlet types, respectively.

477   Finally, to provide a better sense of the nature of the outlets that frequently rely on preprints, we

478 identified the 25 media outlets whose coverage included the largest share of research mentions in

479 general (i.e., mentions of preprints and WoS outputs) and calculated their share of preprint mentions

480 both before and during the COVID-19 era (Table 3). The list represents about 75% of all research

481 mentions in our sample and includes a mix of legacy media, such as *BBC News* and *The New York*

482 *Times*, and peripheral outlets, such as *Reason* or *The Conversation*. Several non-journalism outlets also

483 appear on the list, mostly services such as *EurekAlert!* and *Newswise,* which do not publish original

484 articles but distribute science press releases (many of which include mentions of new research). Among

485 outlets that tended to cover a high proportion of preprints in general, the US libertarian magazine

486 *Reason* stood out, mentioning approximately one preprint for every three WoS outputs—far more than

487 any other outlet in our sample prior to the COVID-19 pandemic. Interestingly, the outlet's share of

488 mentions actually decreased slightly during the pandemic, from 27% to 24%. Among the outlets that

489 saw the largest increase in their share of preprint mentions, the peripheral Health/Medicine outlet *News*

490 *Medical* topped the list, with essentially no preprint mentions before the pandemic but a share of 43%

491 during the pandemic. Several major legacy General News outlets, such as *BBC News, The Daily Mail,*

25

492 *The New York Times,* and *The Guardian,* also saw notable increases in preprint coverage, moving from

493 minimal use of preprints to covering about one preprint for every four or five mentions of research.

494 Although some specialized Science/Technology outlets (e.g., *Scientific American, Phys.org*) increased

495 their coverage of preprints during COVID-19, these increases tended to be less pronounced than those

496 seen among the major General News outlets.

497

498 **Table 3.** Largest 25 media outlets based on mentions and the share of mentions that include preprints

	Outlet's Share of All Research Mentions	Outlet's Share of Preprint Mentions	
		Pre-pandemic Era	**Pandemic Era**
BBC News	1.75%	1.69%	21.85%
Business Insider	2.51%	2.61%	17.45%
Business Insider Australia	1.37%	2.03%	16.32%
Daily Mail	2.07%	1.45%	22.24%
EurekAlert!	1.55%	0.24%	2.83%
Forbes	6.37%	9.94%	12.37%
Gizmodo	1.00%	3.33%	14.83%
MedicalXpress	2.42%	0.19%	7.75%
New Scientist	1.09%	4.15%	20.84%
New York Times	8.12%	3.69%	23.32%
Newswise	1.28%	0.60%	6.98%
Phys.org	3.86%	0.99%	5.63%
Quartz	1.47%	4.93%	16.38%
Reason	2.03%	26.83%	24.01%
Salon	1.25%	4.27%	17.27%
Science/AAAS	0.98%	2.85%	33.62%
Scientific American	1.02%	2.13%	17.17%
The Atlantic	1.57%	6.48%	23.82%
The Conversation	6.84%	1.75%	9.76%
The Guardian	1.88%	2.01%	20.56%
Medical News	8.73%	0.30%	42.70%
Times of India	1.03%	0.99%	11.96%
Vox	1.60%	4.79%	19.23%
Washington Post	3.16%	4.66%	13.68%
Yahoo! News	11.44%	1.57%	14.27%

NB: Top 25 outlets shown based on their share of mentions of Web of Science outputs and preprints, representing about 75% of all mentions in our sample. Column 2 (Outlet's Share of All Research Mentions) shows each outlet's share of all mentions of Web of Science outputs and preprints (i.e., number of preprints and WoS research for the outlet divided by total number of preprints and WoS mentions in our sample). Columns 3 and 4 (Outlet's Share of Preprint Mentions) shows the share of each outlet's mentions of preprints (i.e., number of preprint mentions for the outlet divided by total number of preprint mentions in our sample) prior to COVID-19 and the share of each outlet's mentions for preprints during COVID-19. Grayscale conditional formatting is based on column 2 alone and then columns 3 and 4 jointly. Columns do not sum to 100% because only the top 25 outlets are shown.

27

499 5.0 Discussion

500 It has been argued that preprint coverage during the pandemic constituted a break from journalism

501 norms and a paradigm shift in how emergent research is reported on and shared with the public (Burke,

502 2021; Makri, 2021). Using longitudinal data from the Web of Science (WoS) and four preprint servers,

503 this study sought to establish whether, in what ways, and to what extent this is the case. By identifying

504 how the volume and nature of preprint media coverage has changed over time and what role the

505 pandemic has played in this change, our study makes an important contribution to our understanding of

506 journalists' use of preprints—a topic about which much has been written, but very little is actually

507 known.

508   A key finding from our analysis is that the volume of preprint media coverage increased by

509 roughly fourfold in the pandemic period, a clear break from the slight but steady upward trend that

510 preceded it. Virtually all of this increase was driven by coverage of COVID-19-related preprints, with

511 little change in coverage of preprints on other topics. Although coverage of peer reviewed research

512 continued to exceed preprint coverage—even during the height of the crisis—the growth in coverage of

513 preprints seen during this period may imply a shift in journalistic norms and practices, including a

514 changing outlook on preliminary, unvetted research and its reporting.

515   At the same time, however, we observed a slight (but nonsignificant) decrease in coverage of

516 non-COVID-19-related preprints during the pandemic. This lack of coverage of non-COVID-19-related

517 preprints may simply be the result of outsized media attention to COVID-19 in general (i.e., not just

518 COVID-19-related preprints), which may have come at the expense of coverage on other topics. Yet, it

519 could also indicate that the surge in preprint coverage observed during the pandemic was a temporary

520 change—a break from established norms that journalists made to cover a rapidly evolving crisis, rather

521 than a true shift in practice. More research is needed to assess the degree to which increases in preprint

522   coverage will persist in the coming years, as media outlets and scientists turn their attention away from

523   COVID-19 and toward other issues.

524   Interestingly, the sharp rise in preprint coverage seen during the pandemic was most pronounced

525   for health and medical outlets, which appear to have been resistant to covering preprints until relatively

526   recently. While outlets specializing in other topics, such as science and technology, covered preprints at

527   least occasionally before the pandemic, our findings suggest that, for health and medical outlets, the

528   crisis seems to have created something closer to the "paradigm shift" described by journalists in

529   previous research (Fleerackers et al., 2022a). Preprints were barely mentioned in health and medical

530   outlets up until 2019—even after medRxiv was launched—but become a frequent source of coverage in

531   these outlets after 2020, particularly when reporting on COVID-19. Again, more research is needed to

532   assess whether this trend will continue beyond the pandemic.

533   The factors that motivated health and medical journalists to adapt their practices during COVID-

534   19 also remain unclear. While the medical nature of the crisis likely played a primary role, at least some

535   of this shift may be linked to a parallel shift in preprint use among health and medical scholars

536   themselves. Like journalists, researchers in these areas have historically been hesitant to post or cite

537   unreviewed research (Flanagin et al., 2020; Maslove, 2018), but became active users of preprints during

538   the pandemic (Fraser et al., 2021; Waltman et al., 2021). Since journalists who report on research rely

539   heavily on interviews with scientific experts (Schultz, 2023), changing attitudes toward preprints among

540   medical scientists would likely affect reporting practices on medical and related issues. It is possible, in

541   other words, that the uptake of preprints by medical and health outlets reflects the growing acceptance of

542   preprints within the medical and health sciences. This may also be true of preprint-based journalism

543   more broadly, as preprint adoption also grew during the study period (Nane et al., 2023). Waltman et al.

544   (2021) report that the number of preprints in 2020 was about 150% larger than the number of preprints

545   in 2015, while Penfold and Polka (2020)—working with data from PubMed and 10 preprint servers—

546 found that the number of biology preprints increased almost tenfold between 2013 and 2019 (from

547 0.24% to 2.36%). However, just because more preprints are becoming available doesn't mean journalists

548 will automatically cover them. By covering preprint science, journalists may—potentially—be adapting

549 their own norms to follow those of scientists working in the disciplines they report on. It is also possible

550 that journalistic norms in terms of preprint coverage are changing as journalists are increasingly

551 pressured to attract reader attention away from competing outlets. In other words, the rapid and

552 competitive nature of the media landscape may encourage journalists to use preprints even outside

553 the pandemic context—and thus get an 'edge' over other media outlets, as suggested in previous

554 research (Fleerackers et al., 2022a).

555 In terms of outlet types, we found that both traditional, legacy outlets (e.g., *The New York Times*)

556 and peripheral media outlets (e.g., *News Medical*) were covering preprints to some extent before the

557 pandemic, but greatly increased this coverage during the crisis. The similar pattern seen for the two

558 outlet types is surprising, as peripheral media outlets are often conceptualized as following different

559 norms, ethics, and practices than legacy media and as being less beholden to professional guidelines,

560 such as those that urge journalists to avoid covering unreviewed research (Froke et al., 2020). Our

561 findings thus align with previous scholarship which has suggested that the boundaries between legacy

562 and peripheral journalism are blurring and that categorizing outlets this way may no longer be

563 meaningful (Deuze & Witschge, 2018; Witschge et al., 2019). While more research is needed, it is

564 possible that such blurring boundaries are especially likely in contexts where professional guidelines and

565 best practices are still emerging, such as when reporting on preprints (van Schalkwyk & Dudek, 2022a).

566 Future studies could explore whether the similarities we observed in preprint coverage among peripheral

567 and legacy outlets also apply to larger and more diverse outlet samples, or to other situations in which

568 journalistic practices are rapidly evolving.

569   Collectively, our findings provide some of the first evidence that journalists are increasingly

570   using preprints—at least in some areas—and that the pandemic has greatly accelerated this use.

571   However, this conclusion should be considered alongside several limitations. First, there are known

572   challenges of working with Altmetric data to identify media coverage of research, particularly in

573   languages other than English (see Ortega, 2020a, for a review). We have attempted to mitigate these

574   challenges by working with a predefined set of English-language media outlets, as recommended in

575   previous research (Fleerackers et al., 2022b). Yet, while the restricted nature of our sample of outlets is a

576   strength of this study, it is also a limitation, as the patterns we observed among these 94 outlets may not

577   apply to those that less frequently report on research or do so in different languages. Relatedly, an

578   increasing number of regional preprint servers are coming online, which offer spaces for authors to share

579   preprints that are relevant to their own geographic context, in languages other than English (e.g.,

580   SciELO preprints, AfricArXiv, Jxiv). These servers may become preferred sources for media outlets

581   publishing non-English content. Replicating our findings with a larger set of outlets and servers, or

582   through complementary data gathering methods, would be a fruitful avenue for future research.

583   We also aimed to make our findings more robust by contextualizing any increases in preprint

584   media coverage alongside changes in coverage of peer reviewed research during the same time period.

585   To do so, we relied on Web of Science data, which is biased towards studies from scientific, technical,

586   and medical disciplines and published in English-language journals from the Global North (Alperin et

587   al., 2014; Mongeon & Paul-Hus, 2016). Given our study's focus on English-language media outlets, the

588   impact of the language bias is likely minimal (i.e., it is relatively unlikely that a journalist working for

589   an English-language outlet would cover non-English research). However, the disciplinary and

590   geographic biases are limitations of our study that should be kept in mind when interpreting the results.

591   Finally, the nature of our data only enabled us to explore changes in preprint media coverage

592   from 2014 through the first year and a half of the pandemic, leaving many questions unanswered about

31

593   what the future will bring. We hope that scholars will build on our findings to provide further insight

594   into the implications of the preprint coverage seen during COVID-19 will persist long-term.

Acknowledgements

Funding

Competing Interests

603   None to declare.

Data Availability

605   Preprint and WoS mention data for this study can be accessed via the Harvard Dataverse

606   at https://dataverse.harvard.edu/ with the doi: .7910/DVN/ZHQUFD. Scripts used to analyze the data

607   can be accessed via Zenodo at https://zenodo.org with the doi: NODO.8125008.

References

609   Alperin, J. P., Babini, D., & Fischman, G. (Eds.). (2014). *Open access indicators and scholarly*

610   *communications in Latin America*. CLACSO. http://eprints.rclis.org/25122/

32

611   Alperin, J. P., Fleerackers, & Shores. (2023). *Data for: Unreviewed science in the news* (DRAFT

612   VERSION) [dataset]. Harvard Dataverse. https://doi.org/10.7910/DVN/ZHQUFD

613   Alperin, J. P., Shores, K., & Fleerackers, A. (2023). *ScholCommLab/preprints-over-time: Initial Release*

614   (1.0) [Computer software]. Zenodo. https://doi.org/10.5281/ZENODO.8125008

615   Besançon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M.,

616   & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. *BMC*

617   *Medical Research Methodology*, *21*(1), 117. https://doi.org/10.1186/s12874-021-01304-y

618   Broer, I. (2020). Rapid reaction: Ethnographic insights into the Science Media Center and its response to

619   the COVID-19 outbreak. *Journal of Science Communication*, *19*(5), A08.

620   https://doi.org/10.22323/2.19050208

621   Broer, I., & Pröschel, L. (2022). Knowledge broker, trust broker, value broker: The roles of the Science

622   Media Center during the COVID-19 pandemic. *Studies in Communication Sciences*, *22*(1),

623   Article 1. https://doi.org/10.24434/j.scoms.2022.01.3070

624   Burke, L. (2021, January 20). Pivotal Year for Preprints. *Inside Higher Ed*.

625   https://www.insidehighered.com/news/2021/01/20/pandemic-brought-new-attention-preprints

626   Chamberlain, S. (2020). *Habanero [source code]* (0.7.4) [Computer software]. Github.

627   https://github.com/sckott/habanero/releases/tag/v0.7.4

628   Chen, Q., Allot, A., & Lu, Z. (2020). Keep up with the latest coronavirus research. *Nature*, *579*(7798),

629   193–193. https://doi.org/10.1038/d41586-020-00694-1

630   Deuze, M., & Witschge, T. (2018). Beyond journalism: Theorizing the transformation of journalism.

631   *Journalism: Theory, Practice & Criticism*, *19*(2), 165–181.

632   https://doi.org/10.1177/1464884916688550

633   Else, H. (2020). How a torrent of COVID science changed research publishing—In seven charts.

634   *Nature*, *588*(7839), 553–553. https://doi.org/10.1038/d41586-020-03564-y

33

635   Flanagin, A., Fontanarosa, P. B., & Bauchner, H. (2020). Preprints Involving Medical Research—Do the

636   Benefits Outweigh the Challenges? *JAMA*, *324*(18), 1840.

637   https://doi.org/10.1001/jama.2020.20674

638   Fleerackers, A., Chtena, N., Pinfield, S., Alperin, J. P., Barata, G., Oliveira, M., & Peters, I. (2023).

639   Making science public: A review of journalists' use of Open Science research [version 1; peer

640   review: 1 approved]. In *F1000Research* (Vol. 12, Issue 512).

641   https://doi.org/10.12688/f1000research.133710.1

642   Fleerackers, A., & Fagan, K. (2022). *Analyzing the nature of media outlets* (Version 1) [Computer

643   software]. OSF. https://doi.org/10.17605/OSF.IO/NW87U

644   Fleerackers, A., Moorhead, L. L., Maggio, L. A., Fagan, K., & Alperin, J. P. (2022). Science in motion:

645   A qualitative analysis of journalists' use and perception of preprints. *PLOS ONE*, *17*(11),

646   e0277769. https://doi.org/10.1101/2022.02.03.479041

647   Fleerackers, A., Nehring, L., Maggio, L. A., Enkhbayar, A., Moorhead, L., & Alperin, J. P. (2022).

648   Identifying science in the news: An assessment of the precision and recall of Altmetric.com news

649   mention data. *Scientometrics*, *127*(11), 6109–6123. https://doi.org/10.1007/s11192-022-04510-7

650   Fleerackers, A., Riedlinger, M., Moorhead, L., Ahmed, R., & Alperin, J. P. (2021). Communicating

651   scientific uncertainty in an age of COVID-19: An investigation into the use of preprints by

652   digital media outlets. *Health Communication*, *37*(6), 726–738.

653   https://doi.org/10.1080/10410236.2020.1864892

654   Fraser, N., Brierley, L., Dey, G., Polka, J. K., Pálfy, M., Nanni, F., & Coates, J. A. (2021). The evolving

655   role of preprints in the dissemination of COVID-19 research and their impact on the science

656   communication landscape. *PLOS Biology*, *19*(4), e3000959.

657   https://doi.org/10.1371/journal.pbio.3000959

34

658   Fraser, N., Momeni, F., Mayr, P., & Peters, I. (2020). The relationship between bioRxiv preprints,

659   citations and altmetrics. *Quantitative Science Studies*, *1*(2), 618–638.

660   https://doi.org/10.1162/qss_a_00043

661   Froke, P., Bratton, A. J., McMillan, J., Sarkar, P., Schwartz, J., Vadarevu, R., & Associated Press.

662   (2020). Health, science and environment reporting. In *The Associated Press stylebook 2020-2022*

663   (55th ed.).

664   Funk, K. (2023, February 8). NIH preprint pilot expands to include preprints across NIH-funded

665   research. *NLM Musings from the Mezzanine*. https://nlmdirector.nlm.nih.gov/2023/02/08/nih-

666   preprint-pilot-expands-to-include-preprints-across-nih-funded-research/

667   Haelle, T. (2020, April 17). Beware the preprint in covering coronavirus research. *Association of Health*

668   *Care Journalists*. https://healthjournalism.org/blog/2020/04/beware-the-preprint-in-covering-

669   coronavirus-research/

670   Harrison, S., Macmillan, A., & Rudd, C. (2020). Framing climate change and health: New Zealand's

671   online news media. *Health Promotion International*, *35*(6), 1320–1330.

672   https://doi.org/10.1093/heapro/daz130

673   Haustein, S., Costas, R., & Larivière, V. (2015). Characterizing social media metrics of scholarly

674   papers: The effect of document properties and collaboration patterns. *PLOS ONE*, *10*(3),

675   e0120495. https://doi.org/10.1371/journal.pone.0120495

676   Hermida, A. (2019). The existential predicament when journalism moves beyond journalism.

677   *Journalism*, *20*(1), 177–180. https://doi.org/10.1177/1464884918807367

678   Hermida, A., & Young, M. L. (2019). From peripheral to integral? A digital-born journalism not for

679   profit in a time of crises. *Media and Communication*, *7*(4), Article 4.

35

680 Horbach, S. P. J. M. (2020). Pandemic publishing: Medical journals strongly speed up their publication

681 process for COVID-19. *Quantitative Science Studies*, *1*(3), 1056–1067.

682 https://doi.org/10.1162/qss_a_00076

683 Hurley, R. J., & Tewksbury, D. (2012). News aggregation and content differences in online cancer

684 news. *Journal of Broadcasting & Electronic Media*, *56*(1), 132–149.

685 https://doi.org/10.1080/08838151.2011.648681

686 Jung, Y. E. (Grace), Sun, Y., & Schluger, N. W. (2021). Effect and reach of medical articles posted on

687 preprint servers during the COVID-19 pandemic. *JAMA Internal Medicine*, *181*(3), 395.

688 https://doi.org/10.1001/jamainternmed.2020.6629

689 Kaiser, J. (2019, June 5). Medical preprint server debuts. *Science Magazine*.

690 https://www.science.org/content/article/medical-preprint-server-debuts

691 Kousha, K., & Thelwall, M. (2020). COVID-19 publications: Database coverage, citations, readers,

692 tweets, news, Facebook walls, Reddit posts. *Quantitative Science Studies*, *1*(3), 1068–1091.

693 https://doi.org/10.1162/qss_a_00066

694 Krippendorff, K. (2004). *Content analysis: An introduction to its methodology* (2nd ed). Sage.

695 Kwon, D. (2021, September 10). A Surge in Pandemic Research Shines a Spotlight on Preprints. *The*

696 *Scientist Magazine*. https://www.the-scientist.com/news-opinion/a-surge-in-pandemic-research-

697 shines-a-spotlight-on-preprints-69170

698 Maggio, L., Alperin, J. P., Moorhead, L., & Willinsky, J. (2017, March 21). Can your doctor see the

699 cancer research reported in the news? Can you? *Medium*.

700 https://medium.com/@lauren.maggio01/can-your-doctor-see-the-cancer-research-reported-in-

701 the-news-can-you-beb9270c301f

702 Makri, A. (2021). What do journalists say about covering science during the COVID-19 pandemic?

703 *Nature Medicine*, *27*(1), 17–20. https://doi.org/10.1038/s41591-020-01207-3

36

704 Maslove, D. M. (2018). Medical preprints—A debate worth having. *JAMA*, *319*(5), 443.

705   https://doi.org/10.1001/jama.2017.17566

706 Massarani, L., & Neves, L. F. F. (2021). Reporting COVID-19 preprints: Fast science in newspapers in

707   the United States, the United Kingdom and Brazil. *Cien Saude Colet [Periódico Na Internet]*.

708   http://www.cienciaesaudecoletiva.com.br/artigos/reporting-covid19-preprints-fast-science-in-

709   newspapers-in-the-united-states-the-united-kingdom-and-brazil/18235?id=18235

710 Massarani, L., Neves, L. F. F., Entradas, M., Lougheed, T., & Bauer, M. W. (2021). Perceptions of the

711   impact of the COVID-19 pandemic on the work of science journalists: Global perspectives.

712   *Journal of Science Communication*, *20*(07), A06. https://doi.org/10.22323/2.20070206

713 Mongeon, P., & Paul-Hus, A. (2016). The journal coverage of Web of Science and Scopus: A

714   comparative analysis. *Scientometrics*, *106*(1), 213–228. https://doi.org/10.1007/s11192-015-

715   1765-5

716 Nane, G. F., Robinson-Garcia, N., van Schalkwyk, F., & Torres-Salinas, D. (2023). COVID-19 and the

717   scientific publishing system: Growth, open access and scientific fields. *Scientometrics*, *128*, 345–

718   362. https://doi.org/10.1007/s11192-022-04536-x

719 Ortega, J. L. (2020a). Altmetrics data providers: A meta-analysis review of the coverage of metrics and

720   publication. *El Profesional de La Información*, *29*(1). https://doi.org/10.3145/epi.2020.ene.07

721 Ortega, J. L. (2020b). Blogs and news sources coverage in altmetrics data providers: A comparative

722   analysis by country, language, and subject. *Scientometrics*, *122*(1), 555–572.

723   https://doi.org/10.1007/s11192-019-03299-2

724 Penfold, N. C., & Polka, J. K. (2020). Technical and social issues influencing the adoption of preprints

725   in the life sciences. *PLOS Genetics*, *16*(4), e1008565.

726   https://doi.org/10.1371/journal.pgen.1008565

37

727 Puebla, I., Polka, J., & Rieger, O. (2021). *Preprints: Their Evolving Role in Science Communication*

728   [Preprint]. MetaArXiv. https://doi.org/10.31222/osf.io/ezfsk

729 Puebla, I., Polka, J., & Rieger, O. Y. (2022). *Preprints: Their evolving role in science communication*.

730   Against the Grain (Media), LLC. https://doi.org/10.3998/mpub.12412508

731 Rosen, C., Guenther, L., & Froehlich, K. (2016). The question of newsworthiness: A cross-comparison

732   among science journalists' selection criteria in Argentina, France, and Germany. *Science

733   Communication*, *38*(3), 328–355. https://doi.org/10.1177/1075547016645585

734 Saari, M.-A., Gibson, C., & Osler, A. (1998). Endangered species: Science writers in the Canadian daily

735   press. *Public Understanding of Science*, *7*(1), 61–81.

736   https://doi.org/10.1177/096366259800700105

737 Schäfer, M. S. (2017). The Oxford handbook of the science of science communication. In K. H.

738   Jamieson, D. M. Kahan, & D. A. Scheufele (Eds.), *How Changing Media Structures Are

739   Affecting Science News Coverage* (Vol. 1). Oxford University Press.

740   https://doi.org/10.1093/oxfordhb/9780190497620.013.5

741 Schapals, A. K. (2022). *Peripheral actors in journalism: Deviating from the norm?* Routledge (Taylor &

742   Francis). https://doi.org/10.4324/9781003144663

743 Schultz, T. (2023). A survey of U.S. science journalists' knowledge and opinions of open access

744   research. *International Journal of Communication*, *17*, Article 0.

745 Schwab, L. (2021). *arxiv: Python wrapper for the arXiv API: http://arxiv.org/help/api/* (1.4.2) [Python;

746   OS Independent]. 0.7.4. https://github.com/lukasschwab/arxiv.py

747 Shoemaker, P. J., & Reese, S. D. (1996). *Mediating the message: Theories of influences on mass media

748   content* (2nd ed). Longman.

749 Simons, A., & Schniedermann, A. (2023). Preprints in the German news media before and during the

750   COVID pandemic. A comparative mixed-method analysis. In I. Broer, S. Lemke, A. Mazarakis,

38

751   I. Peters, & C. Zinke-Wehlmann (Eds.), *The Science-Media Interface: On the Relation Between*

752   *Internal and External Science Communication* (pp. 53–77). De Gruyter Saur.

753   https://www.degruyter.com/document/isbn/9783110776546/html?lang=en

754 StataCorp. (2021). *Stata Statistical Software* (Version 17) [Computer software]. StrataCorp LLC.

755 Stocking, G. (2019). *Trends and facts on online news: State of the news media*. Pew Research Center.

756   https://www.journalism.org/fact-sheet/digital-news/

757 The Pandas Development Team. (2023). *Pandas* (1.5.3) [Python].

758   https://doi.org/10.5281/zenodo.7549438

759 van Schalkwyk, F., & Dudek, J. (2022a). Reporting preprints in the media during the Covid-19

760   pandemic: Supplemental material. *Public Understanding of Science*, 2.

761 van Schalkwyk, F., & Dudek, J. A. (2022b). *Reporting uncertified science in the news media during the*

762   *Covid-19 pandemic*. AfricArXiv. https://doi.org/10.14293/111.000/000025.v1

763 Vergoulis, T., Kanellos, I., Chatzopoulos, S., Pla Karidi, D., & Dalamagas, T. (2021). BIP4COVID19:

764   Releasing impact measures for articles relevant to COVID-19. *Quantitative Science Studies*,

765   *2*(4), 1447–1465. https://doi.org/10.1162/qss_a_00169

766 Waltman, L., Pinfield, S., Rzayeva, N., Henriques, S. O., Fang, Z., Brumberg, J., Greaves, S., Hurst, P.,

767   Collings, A., Heinrichs, A., Lindsay, N., MacCallum, C., Morgan, D., Sansone, S.-A., &

768   Swaminathan, S. (2021). *Scholarly communication in times of crisis* [Report]. Research on

769   Research Institute. https://apo.org.au/node/315479

770 WHO. (2020, March 19). *Infection prevention and control during health care when novel coronavirus*

771   *(nCoV) infection is suspected*. WHO. https://www.who.int/publications-detail-redirect/10665-

772   331495

773 Wihbey, J. (2017). Journalists' use of knowledge in an online world: Examining reporting habits,

774   sourcing practices and institutional norms. *Journalism Practice*, *11*(10), 1267–1282.

775   https://doi.org/10.1080/17512786.2016.1249004

776 Witschge, T., Anderson, C., Domingo, D., & Hermida, A. (2019). Dealing with the mess (we made):

777   Unraveling hybridity, normativity, and complexity in journalism studies. *Journalism*, *20*(5),

778   651–659. https://doi.org/10.1177/1464884918760669

779