

1 Title

2 Full Publication of Preprint Articles in Prevention Research: An Analysis of Publication

3 Proportions and Results Consistency

4 Short title

5 Publication Proportions and Results Consistency of Preprint Articles in Prevention Research

6 Authors

7 Isolde Sommer¹, Vincent Sunder-Plassmann¹, Piotr Ratajczak², Robert Emprechtinger³, Andreea

8 Dobrescu¹, Ursula Griebler¹, Gerald Gartlehner^{1,4}

9 Affiliations

10 ¹Cochrane Austria, Department for Evidence-based Medicine and Evaluation, Danube University Krems,
11 Krems, Austria

12 ²Department of Pharmacoeconomics and Social Pharmacy, Poznań University of Medical Sciences,
13 Poznań, Poland

14 ³Faculty of Health and Medicine, Danube University Krems, Krems, Austria

15 ⁴RTI International, Research Triangle Park, North Carolina USA

16 Corresponding author:

17 Isolde Sommer

18 Email: isolde.sommer@donau-uni.ac.at

19

20 Word Count:

21 **Number of tables: 3**

22 **Number of figures: 1**

23 Abstract

24 Introduction

25 There is concern that preprint articles will lead to an increase in the amount of scientifically invalid work
26 available. The objectives of this study were 1) to determine the proportion of prevention preprints
27 published within 12 months, 2) to assess the consistency of the effect estimates and conclusions
28 between preprint and published articles, and 3) to explore the reasons for the nonpublication of
29 preprints.

30 Methods

31 We developed a web crawler to search the preprint server medRxiv for prevention studies posted from
32 January 1 to September 30, 2020. An update search was conducted 12 months later. We dually screened
33 the results for prevention articles and developed a scheme to classify changes in effect sizes and
34 conclusions. We modeled the effect of a set of predictors on the proportion of preprints published in
35 peer-reviewed journals. We also developed a questionnaire for authors of unpublished preprints.

36 Results

37 Of the 329 prevention preprints that met our eligibility criteria, almost half (48.9%) were published in a
38 peer-reviewed journal within 12 months of being posted, with the median time being 5.3 months (range
39 0.1–11.3 months). While 27 out of 161 (16.8%) published preprints showed some change in the
40 magnitude of the primary outcome effect estimate, 4.4% were classified as having a major change. The
41 style or wording of the conclusion changed in 42.2%, while the content of the conclusion changed in
42 3.1%. Preprints on chemoprevention, with a cross-sectional design, and with public and noncommercial

43 funding had the highest probabilities of publication. The main reasons for the nonpublication of
44 preprints were journal rejection or lack of time.

45 **Conclusion**

46 The reliability of preprint articles for evidence-based decision-making is questionable. Less than half of
47 the preprint articles on prevention research are published in a peer-reviewed journal within 12 months,
48 and significant changes in effect sizes and/or conclusions are still possible during the peer-review
49 process.

50

51 Introduction

52 Preprints in health sciences have a relatively short tradition compared to other fields (e.g., physics,
53 mathematics, biology) where researchers have been using preprint servers since the 1990s to distribute
54 their research findings and ideas (1). With the founding of the medical preprint server medRxiv
55 (www.medrxiv.org) in June 2019, preprints entered the field of medical and health research. The server's
56 popularity dramatically increased during the coronavirus disease 2019 (COVID-19) pandemic, which has
57 boosted its use (2). These days, many journals (e.g., Elsevier [3], Springer [4], PLOS ONE [5], Lancet [6])
58 have introduced a preprint policy that allows or even encourages the sharing of preprints prior to peer-
59 reviewed publication. Funding agencies such as Wellcome permit researchers to cite preprints in grant
60 applications (7), and, in addition to encouraging preprint postings, they even require it if preprints being
61 shared widely and rapidly results in a significant public health benefit (8).

62 Indeed, early and fast dissemination is the most appealing feature of preprints (9). Quick research
63 sharing enables other researchers to build on early results, accelerating the research efforts necessary to
64 overcome pressing health issues (10). There are concerns, however, that circumventing the peer-review
65 process leads to an increase in the amount of scientifically invalid work (9). Some preprints have been
66 cited widely in the press (10) and, without communicating the proper caution, there is a risk that their
67 findings can be exaggerated by the media, while better-quality work could be ignored (11). According to
68 a study in South Africa, 59% of news articles citing preprints failed to provide a statement of
69 provisionality (12).

70 Various publications have discussed the validity of preprints, but despite the increasing amount of
71 evidence on reporting and quality assurance, the available knowledge is still restricted to COVID-19,
72 biomedical, or interdisciplinary research (13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23). No information exists
73 on the use and validity of preprints in prevention research. Evidence from prevention research impacts
74 community health and public health practice and informs public and policy decision-making every day,

75 not only during emergent public health crises. It is therefore crucial to understand the validity of preprint
76 results and conclusions in prevention research. The objectives of this study were

77 1) to determine the proportion of preprints that are published within 12 months of being added to
78 medRxiv, overall and between different prevention types,

79 2) to assess the consistency of the effect estimates and conclusions between the preprint and
80 published versions of prevention articles, and

81 3) to explore the reasons for the nonpublication of preprints in peer-reviewed journals.

82 **Materials and Methods**

83 Our study protocol was registered in the Open Science Framework (OSF) (24). The overall project was a
84 mixed-methods study and consisted of three parts: a text analysis, a qualitative interview study, and a
85 survey study. Here, we report the results of the text analysis (i.e., objectives 1 and 2 of the larger overall
86 project) as well as the results of the survey study, which we registered as an update to the original
87 protocol in OSF (25).

88 **Text analysis**

89 ***Data source and search strategy***

90 We sampled studies from medRxiv (www.medrxiv.org). We developed a Python-based web crawler by
91 analyzing the medRxiv website's http responses using the Python packages requests (handling http) (26),
92 BeautifulSoup (xml/html parsing) (27), and re (for extracting information from character strings related
93 to prevention using regular expressions) (28) (see Table S1 for search string). We ran the web crawler on
94 December 15, 2020 and let it search and extract information from prevention articles that had been
95 posted on the medRxiv website (first run) from January 1, 2020 to September 30, 2020. It downloaded
96 basic data about each identified preprint article: title, abstract, authors, version submission date, version

97 history, download statistics, withdrawal information, funder, first author's institutional affiliation, and
98 information on the publication status (if published, new DOI, and the journal in which it appeared).

99 ***Inclusion and exclusion criteria for the study selection***

100 We exported the information provided by the web crawler to Excel® for the abstract review. We dually
101 screened all records identified by the web crawler against our eligibility criteria (see Table 1). Within the
102 project, we used the working definition for prevention research established by the National Institute of
103 Health Prevention Research (NIHR) Coordinating Committee (29). Using that definition, we included
104 primary and secondary prevention studies that (1) identified and assessed risk and protective factors, (2)
105 screened and identified individuals and groups at risk, (3) developed and evaluated interventions to
106 reduce risk, (4) translated, implemented, and disseminated effective preventive interventions into
107 practice, or (5) developed methods to support prevention research. We pilot-tested the abstract review
108 with 50 records in the first web crawler round and amended the eligibility criteria where necessary. In
109 case of uncertainty, we looked at the preprint's full text and solved disagreements through discussion.
110 We dually categorized each record according to the prevention categories (i.e., chemoprevention,
111 counseling, immunization, screening, other primary prevention, other secondary prevention), whether
112 COVID-19-related (i.e., yes or no), funding source (i.e., any funding vs. no funding, public or
113 noncommercial funding, public and noncommercial funding, industrial funding, no funding), and study
114 design (i.e., randomized controlled trial [RCT], cohort study, cross-sectional study, diagnostic study,
115 ecological study, descriptive study, time series, before-after study, case control study, case series).

116 Table 1: Inclusion and exclusion criteria for the selection of preprints to be included in the text analysis

	Inclusion criteria	Exclusion criteria
Topic	All prevention research (see the National Institute of Health Prevention Research [NIHR] Coordinating Committee's definition) (29), which can be categorized into: chemoprevention, counseling, immunization, screening, other primary prevention, other secondary prevention	Tertiary prevention, treatments, all other topics
Study designs	Clinical studies (including phase 2	Modeling studies based on nonreal

	trials), epidemiological studies, diagnostic studies	data, in vitro studies, qualitative studies, cost-effectiveness studies, all reviews (also systematic and rapid reviews), and basic research studies
Publication status	Preprints	All other publication types
Language	English	Other than English
Date posted on medRxiv	January 1, 2020 to September 30, 2020	Later than September 30, 2020

117

118 ***Search for published preprint articles***

119 To give every preprint a 12-month span to get published, we ran the web crawler again on October 5,
120 2021 and updated the information on publication status (second run). Because we did not want to rely
121 entirely on the information provided by medRxiv regarding publication status, we manually searched
122 Google® and GoogleScholar® for a published version of each unpublished preprint. If we still failed to
123 identify a published version, we contacted the corresponding author of the unpublished preprint by
124 email.

125 ***Data extraction and analysis***

126 For preprints that were published in a peer-reviewed journal, we downloaded the article and performed
127 further data extractions into a structured form using Excel®. One researcher extracted the following
128 data, which was checked by a second researcher: the primary outcome effect estimate and conclusions
129 regarding the primary outcome for both the preprint and peer-reviewed article, journal name, and
130 publication date.

131 When we detected differences in the effect estimates or conclusions between the preprint and peer-
132 reviewed article, two investigators independently classified these changes. We used the typology
133 developed by Gartlehner et al. (30) to classify these changes but had to adapt it because of the range of
134 effect estimates we identified. We considered the statistical significance of the primary outcome
135 between the preprint and peer-reviewed article as having changed when at least one of the two effect
136 estimates had a P-value that was deemed statistically significant in either the preprint or publication, and

137 not statistically significant in the other. We assessed the magnitude of change in the effect estimate by
138 applying the following categories to both dichotomous and continuous outcomes: no change, minor
139 change (a relative change of up to 25 percentage points), and major change (a relative change of more
140 than 25 percentage points).

141 For changes in the conclusion, we followed the categories suggested by Silagy et al. (31): no change,
142 minor change (changes in style or wording that do not alter the substance or meaning of a section), and
143 major change (changes that alter the substance or meaning of a section or alter the interpretation). The
144 classifications of changes in the effect estimates and conclusions were done by one person and verified
145 by a second person.

146 We retrieved the impact factor for each peer-reviewed journal from the Impact Factor List of 2019
147 provided by the Journal Citation Report (JCR) (32) and calculated the time until publication from the first
148 appearance on medRxiv.

149 We used descriptive statistics and compared differences in publication characteristics and publication
150 proportion between the different types of prevention articles. We used Bayesian methods to model the
151 effect of a set of predictors on the proportion of peer-reviewed journal preprint publications as an
152 outcome. The predictors included prevention type, whether COVID-19–related (yes/no), study design,
153 and funding sources. We chose to use multiple individual models instead of one joint model for
154 exploratory reasons, due to the lack of an underlying theory and to avoid issues associated with
155 overadjustment and collider bias (33, 34, 35). The Bayesian modeling was conducted with Markov Chain
156 Monte Carlo methods via the brms package (36) and using restrictive priors. The intercept was
157 suppressed. The statistical models were as follows:

$$Outcome_i \sim \text{Bernoulli}(\mu_i)$$

$$\text{logit}(\mu_i) = \beta_{\text{predictor}[i]}$$

$$\beta_i \sim \text{Normal}(0, 3)$$

158 We conducted all analyses within the R environment (version 4.2.1). Additionally, we used the tidyverse
159 (37) package, readxl (38), and tidybayes (39, 40) packages for data wrangling and creating the plots.

160 **Survey study**

161 ***Survey development, participants, and procedure***

162 We designed a questionnaire to explore the reasons for nonpublication in peer-reviewed journals and
163 attitudes toward preprints in general. The questionnaire was developed in English and consisted of 11
164 items asking about the rationale behind deciding not publish the preprint or the reason(s) for and
165 number of rejections as well as the estimated credibility of preprints, attitudes toward preprints, and
166 demographic characteristics. The questionnaire was developed based on the results from the text
167 analysis. The face validity was confirmed by the research team. Using the correspondence email address
168 provided in the preprint, we sent the questionnaire to all corresponding authors of the prevention
169 preprints identified in the text analysis that were not published at the start of the survey (n=152)
170 between September 14 and November 14, 2022. We sent out several reminder emails to increase the
171 response rate.

172 ***Ethics approval and compliance***

173 The ethics committee of the University for Continuing Education, Krems, approved the survey study (EK
174 GZ 28/2021-2024). Throughout the project, we adhered to the European Union data protection law (EU
175 Regulation 2016/679). Participants were not reimbursed for their participation.

176 **Results**

177 **Characteristics of the included preprints**

178 In the first run, the web crawler identified 2238 preprints on medRxiv, of which we identified 594 as
 179 prevention research studies (26.5%). Among those, 329 were clinical, epidemiological, or diagnostic
 180 studies and met the inclusion criteria for our study selection (Fig. 1).

181 Fig. 1: Flow chart of the preprint selection process

182 Table 2 shows the characteristics of the included preprints. Of all the identified preprints, 73.6%
 183 (242/329) were on a topic related to COVID-19. Almost half focused on screening (46.2%), less than one-
 184 third (27.4%) were on other primary prevention topics, and about one-fifth were on immunization
 185 (20.1%). Few preprints fell into the other prevention categories (0.6% to 4.9%). The proportion of
 186 preprints receiving external funding was highest among the screening and chemoprevention preprints
 187 (59.9% and 62.5%). We identified most preprints as diagnostic (30.4%, 100/329) and cross-sectional
 188 studies (29.8%, 98/329) (Table S2). More than half had received external funding (48.6%, 160/329),
 189 mainly from public or noncommercial funding sources (29.8%, 98/329) (Table S3).

190 Table 2: Characteristics of the included preprints in total and according to prevention type

	Total n/N (%)	COVID-19 topic n/N (%)	External funding n/N (%)
Total	329	242/329 (73.6%)	169/329 (51.4%)
Prevention type			
Screening	152/329 (46.2%)	107/152 (70.4%)	91/152 (59.9%)
Immunization	66/329 (20.1%)	45/65 (69.2%)	30/66 (45.4%)
Chemoprevention	16/329 (4.9%)	8/16 (50.0%)	10/16 (62.5%)
Counseling	2/329 (0.6%)	0/2 (0%)	2/2 (100%)
Other primary prevention	90/329 (27.4%)	81/90 (90.0%)	34/90 (37.8%)
Other secondary prevention	3/329 (0.9%)	1/3 (33.3%)	1/3 (33.3%)

191

192 **Proportion of published preprints**

193 Of the 329 prevention preprints we identified, 161 (48.9%) were published in a peer-reviewed journal
 194 within 12 months of being uploaded to medRxiv (see Table 3). For published preprints, the median time
 195 from upload to journal publication was 5.3 months (range -0.1–11.3). COVID-19 studies were published

196 more quickly than non-COVID studies (4.7 months; range -0.1¹–11.3 vs. 7.0 months; range 1.3–11.3). The
 197 journals those preprints were published in had a median impact factor of 3.2 (range 0.2–74.7).

198 ***Proportions of published preprints according to prevention type, study design, and funding source***

199 The proportions of published preprints differed between the prevention types, with preprints on
 200 screening having the lowest publication proportion (38.2%, 58/152) and those on chemoprevention
 201 having the highest proportion (87.5%, 14/16) (see Table 3). While the time from preprint to peer-
 202 reviewed journal publication showed little difference between prevention categories, the median impact
 203 factors of the journal preprints were highest in the other secondary prevention category, at 7.8 (range
 204 5.4–10.2). For the study design, only about one-third of the diagnostic and ecological studies were
 205 published within 12 months (32.0%, 32/100 and 31.5%, 17/54), while 71.4% of the cross-sectional
 206 studies (70/98) were published within 12 months (see Table S4). The publication process took the least
 207 time for the cohort studies (median: 4.8; range 0.3–11.3) uploaded as preprints as compared to > 6
 208 months for the diagnostic studies. According to the funding source, publicly and noncommercially
 209 funded studies had the highest proportions of published preprints (76.2%, 32/42) and impact factors
 210 (median 5.3, range 0.7–59.1). However, the publication of these studies also took the longest (median
 211 6.2 months; range -0.1–9.8). In contrast, studies receiving no external funding were published faster
 212 (median 2.7 months, range 0.2–3.1) but in journals with lower impact factors (median 4.7, range 0.2–
 213 11.0) (Table S5).

214 Table 3: Proportions of published preprints in total and according to prevention type

	Published within 12 months n/N (%)	Median months until publication (range)	Median impact factor of journal where preprint was published (range)
Total	161/329 (48.9%)	5.3 (-0.1*–11.3)	3.2 (0.2–74.7)
COVID-19	116/242 (47.9%)	4.7 (-0.1*–11.3)	3.2 (0.2–74.7)
Non-COVID-19	45/87 (51.7%)	7.0 (1.3–11.3)	3.4 (0.67–59.1)
Prevention type			
Screening	58/152 (38.2%)	5.9 (-0.1*–11.3)	3.6 (0.2–53.4)
Immunization	31/66 (47.0%)	4.7 (1.0–10.8)	3.6 (0.8–59.1)
Chemoprevention	14/16 (87.5%)	6.6 (0.3–10.6)	3.2 (1.7–74.7)
Counseling	2/2 (100%)	4.7 (4.2–5.3)	3.8 (2.1–5.6)
Other primary	54/90 (60.0%)	4.7 (0.2–10.8)	2.9 (0.5–21.6)

prevention			
Other secondary prevention	2/3 (66.7%)	5.1 (5.0–5.2)	7.8 (5.4–10.2)

215 *Date of publication before first appearance on medRxiv

216 ***Publication predictors***

217 Figures S1–S4 illustrate the effect of prevention type, COVID-19 (yes/no), study design, and funding
218 sources as predictors of preprint publication. Based on our models, a higher probability of preprint
219 publication was associated with chemoprevention topics (88%, 95% Credible Interval [CrI] 67.6 to 97.4),
220 but only few studies were included in these categories. Cross-sectional study preprints (72%, 95% CrI
221 62.1 to 79.6) had the highest publication probability according to the study design. In terms of funding
222 source, studies receiving public and noncommercial funding were the most likely to be published (77%,
223 95% CrI 62.3 to 87.3), while those with no external funding were the least likely (38%, 95% CrI 30.9 to
224 45.7). No difference was observed between COVID-19 and non-COVID-19 prevention studies (52%, 95%
225 CrI 41.7 to 61.8 vs. 48%, 95% CrI 41.8 to 54.1).

226 **Consistency of the effect estimates**

227 Out of 161 preprints that were published in a peer-reviewed journal, 16.8% (27/161) showed a change in
228 the magnitude of the primary outcome effect estimate. This change was major (i.e., greater than 25%) in
229 4.4% studies (7/161) and minor (i.e., less than 25%) in 12.4% studies (20/161). Major changes comprised
230 changes in the effect estimate magnitudes (1.9%, 3/161), in the primary outcomes (1.2%, 2/161), in the
231 type of effect measure (0.6%, 1/161), and in assessment points (0.6%, 1/161). Among the 77 studies
232 reporting statistical significance (47.8%), we did not observe any changes between the preprint and the
233 peer-reviewed journal report, neither from statistically significant to nonsignificant nor vice versa.

234 **Consistency of the conclusions**

235 The conclusions changed in 42.2% (68/161) of the articles after being published in a peer-reviewed
236 journal compared to the preprint, mainly in terms of style or wording (39.1%, 63/161) (i.e., minor
237 change). The content or meaning of the conclusion (i.e., major change) changed in 5 articles (3.1%).

238 **Survey of authors not publishing their preprint**

239 We received a valid answer from 12 out of the 152 authors of preprints not published in a peer-reviewed
240 journal within 12 months and with a valid email address (7.9% response rate). Eleven respondents were
241 male, five were 50 years or older, and eight had more than 10 years of experience in research (Table S6).
242 The reason most often given for nonpublication of the preprint in a peer-reviewed journal was rejection
243 by at least one journal (58.3%, 7/12), followed by lack of time (25%, 3/12). Other reasons mentioned by
244 one respondent each were that the preprint had received its attention, and that they had never intended
245 to publish the preprint. Among those preprints that were submitted, 57.1% (4/7) got rejected 3–4 times.
246 The official reasons given by journals for rejecting the preprints were manifold and included lack of
247 novelty (n=3), too few figures/tables (n=2), and not meeting the journal's scope (n=2), among others.
248 Two-thirds of the preprints did not receive external funding (66.7%, 8/12). The reasons most often
249 indicated for uploading the preprint were sharing the results with the community (n=11), immediate/fast
250 publication (n=6), and increased visibility of work (n=6). All the results are presented in Table S7.

251 **Discussion**

252 **Main findings**

253 To the best of our knowledge, this study is the first to provide a thorough analysis of the publication
254 proportion and consistency of the effect estimates and conclusions of preprints and their subsequent
255 publications in prevention research. Almost half of the prevention preprints (48.9%) were published in a
256 peer-reviewed journal after 12 months of being uploaded to medRxiv, with the median time from upload
257 to publication being 5.3 months (range -0.1*–11.3 months). About half of the prevention preprints were

258 on screening (46.2%), a quarter on other primary prevention topics (27.4%), and one-fifth on
259 immunization (20.1%). The results from the modeling analysis indicate that preprints on
260 chemoprevention and cross-sectional studies or those with public and noncommercial funding had the
261 highest probability of publication within their categories. Preprint authors who did not publish their
262 results in peer-reviewed journals mentioned journal rejections followed by lack of time as the main
263 reasons for nonpublication.

264 We detected a change in the magnitude of the effect estimates in 27 out of the 161 preprints that were
265 published in a peer-reviewed journal (16.8%), but most were minor changes. Although changes in the
266 magnitude of the effect estimate were predominately minor and did not appear very often, they still
267 warrant caution for the use of preprints in decision-making in the prevention field. If 7 out of 161 articles
268 had a major change in the magnitude of the effect estimate, every 23rd article is affected. In addition, it
269 must be considered that as yet we have no knowledge of the quality of unpublished results. We found
270 changes in the conclusions in 42.2% of the preprints that were published within 12 months, but mostly in
271 terms of style or wording, and only in 5 out of 161 articles was the content of the conclusion changed
272 (3.1%). It therefore seems very sensible that medRxiv has issued a warning on the main page of their
273 website that preprints should not be relied on to guide clinical practice or health-related behavior and
274 should not be reported as established information by news media (41). A definite assessment of the
275 credibility of preprints will be possible when the reasons for not publishing them are fully understood.

276 **Relation of the study's findings to previous work**

277 Several studies have centered their work on the publication proportions of preprint articles. Recently,
278 studies focusing on COVID-19 preprints reported proportions of 5.7% to 55.3% preprints published
279 within the study periods (5 to 18 months), with a median time to publication between 2.3 and 5.9
280 months (13, 18, 19, 20). A third of the preprint articles uploaded to bioRxiv, a preprint server for biology
281 research (<https://www.biorxiv.org/>), before 2017 did not get published as peer-reviewed articles (21).

282 The proportion of preprints published in our study (48.9%) is close to that reported in the studies by
283 Otridge et al. (18) and Zeraatkar et al. (19), but these are not directly comparable given the different
284 study periods (12 months vs. 16 and 18 months). Although we did not find that whether an article was
285 COVID-19–related predicted the publication proportion, our findings demonstrate that such articles are
286 published more quickly. An analysis of COVID-19 articles from January to June 2020 showed that peer
287 review was accelerated for COVID-19 articles but decelerated for non-COVID-19 articles because all
288 resources were pushed toward COVID-19 (42). Other studies found more COVID-19 studies published
289 within their study period than non-COVID-19 studies (22, 43).

290 Like in our study, Zeraatkar et al. (19) investigated predictors of preprint publication but came up with
291 different results. They found that preprints were more likely to be published if they received government
292 funding. Our study identified public and noncommercial funding as the strongest predictor for publishing
293 but used different categories for funding source. We further found that chemoprevention and cross-
294 sectional studies had the highest publication probabilities within their categories. To fully understand
295 which factors predict preprint publication, it is important to undertake a larger, more detailed analysis of
296 preprints.

297 As for the changes in effect estimates and conclusions, our study’s findings largely mirror those of other
298 studies reporting on the consistency between preprints and subsequently published articles (14, 15, 16,
299 19). For example, Bero and colleagues (14) did not find large discrepancies in results reporting or the
300 presence of spin between COVID-19 interventional and observational preprints and publications, but
301 small changes were frequent. Another study using a stricter classification scheme (important change in
302 any effect estimate by $\geq 10\%$ and/or change in significance level) classified 21% of COVID-19 intervention
303 trials as having an important change from preprint to peer-reviewed article (44), which is much higher
304 than in our study, as we found that only 4.4% of studies had a major change in the effect estimates.

305 The debate around the credibility of preprints opens the discussion on the quality of peer review. In a
306 study on articles published in bioRxiv and in PubMed-indexed journals, peer-reviewed articles had a
307 higher reporting quality than preprints, particularly in how clearly the titles and abstracts presented the
308 main findings and how easy it was to locate relevant reporting information (45). Changes in reporting
309 from preprints to peer-reviewed versions did not correlate with the journal's impact factor, a finding that
310 resonates with the range of impact factors in our study (median 3.2, range 0.2–74.7). Garcia-Costa et al.
311 (46), who assessed the developmental function of peer review in their study, concluded that taking more
312 time to deliver a peer review increases a manuscript's rigor and value through constructive feedback.
313 Given the increasing challenges journal editors face in recruiting qualified peer-reviewers, the concept of
314 community-based peer review seems appealing. Community-based peer review of posted preprints
315 allows quick scrutiny of research and fast withdrawals when concerns are raised (47). However, it is
316 unclear how widely community-peer review is used and what its impact on the quality of preprints is.

317 **Strengths and limitations of the study**

318 The strengths of this study include the use of a web crawler, which allowed us to automatically screen
319 the medRxiv server, identify preprints in prevention research, and extract relevant information. The web
320 crawler also made it easy to track the publication status of these preprints after 12 months. Another
321 strength is the dual screening, dual data extraction, and dual categorization of the prevention preprints.
322 Finally, we performed a thorough assessment of the included preprint prevention articles, ranging from
323 publication proportion to overall changes in effect size and conclusion and the identification of
324 publication predictors. We are very confident not to have missed any preprint publications within the
325 study period, as we contacted all preprint authors for assurance. Another strength is the additional
326 survey among the preprint authors whose preprints were not published after 12 months to gain further
327 insight on the reasons for nonpublication. A limitation, however, is the low response rate (7.9%), which
328 precluded us from making generalized conclusions.

329 The limitations of the study include the small sample size to analyze differences across the prevention
330 categories, funding sources, and study designs. We further made decisions on the categorization of
331 prevention research, study design, and funding source based only on the information provided in the
332 abstracts; therefore, it is possible that we miscategorized some of them. While the correct classification
333 was important for identifying the predictors, it did not have an influence on the assessment of the
334 changes in the effect estimates and conclusions between the preprint and peer-reviewed article.

335 Conclusions

336 This study expands our knowledge that preprints on prevention research topics have few major changes
337 in the effect estimates and conclusions after undergoing the peer-review process. Although, at first sight,
338 these changes appear in a small number of preprints, still, every 23rd article is affected. We therefore
339 warrant caution in using preprints of prevention research in decision-making.

340 Acknowledgments

341 We would like to thank Florian Nehonsky for developing the web crawler, Irma Klerings for providing the
342 search string, Andrea Trampert for help with the literature screening, Emma Persad for second-checking
343 the effect size and conclusion classifications, and Manuela Müllner for administrative support.

344 Author contributions

345 Conceptualization: IS, GG, UG, AD

346 Investigation: IS, VSP, PR, AD, UG

347 Formal analysis: RE

348 Writing – Original Draft Preparation: IS

349 Writing – Review & Editing: IS, GG, UG, VSP, PR, RE, AD

350 Competing interests

351 The authors declare they have no competing interests in relation to this study.

352 Funding information

353 The study did not receive external funding.

354 Supporting information

355 Table S1: Search string for the web crawler

356 Table S2: Characteristics of the included preprints according to study design

357 Table S3: Characteristics of the included preprints according to funding source

358 Table S4: Proportions of published preprints according to study design

359 Table S5: Proportions of published preprints according to funding source

360 Figure S1: Prevention type as a predictor of preprint publication (probability expressed as %, 95%
361 credible intervals)

362 Figure S2: COVID-19 (yes/no) as a predictor of preprint publication (probability expressed as %, 95%
363 credible intervals)

364 Figure S3: Study design as a predictor for preprint publication (probability expressed as %, 95% credible
365 intervals)

366 Figure S4: Funding source as a predictor for preprint publication (probability expressed as %, 95%
367 credible intervals)

368 Table S6: Sample characteristics of authors of nonpublished preprints (n=12)

369 Table S7: Survey results (n=12)

370 References

- 371 1. Cobb M. The prehistory of biology preprints: A forgotten experiment from the 1960s. *PLOS Biol.*
372 2017;15(11):e2003995.
- 373 2. Krumholz HM, Bloom T, Sever R, Rawlinson C, Inglis JR, Ross JS. Submissions and Downloads of
374 Preprints in the First Year of medRxiv. *JAMA.* 2020;324(18):1903-5.
- 375 3. Article Sharing: Elsevier; 2023 [Available from:
376 <https://www.elsevier.com/about/policies/sharing>.] Accessed May 11, 2023.
- 377 4. Preprint sharing: Springer Nature; 2023 [Available from: [https://www.springer.com/gp/open-](https://www.springer.com/gp/open-access/preprint-sharing/16718886)
378 [access/preprint-sharing/16718886](https://www.springer.com/gp/open-access/preprint-sharing/16718886).] Accessed May 11, 2023.
- 379 5. Preprints: PLOS; 2023 [Available from: <https://journals.plos.org/plosone/s/preprints>.
- 380 6. Kleinert S, Horton R, Editors of the Lancet family of j. Preprints with The Lancet: joining online
381 research discussion platforms. *Lancet.* 2018;391(10139):2482-3.
- 382 7. Open access policy: Wellcome; 2023 [Available from: [https://wellcome.org/grant-](https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy)
383 [funding/guidance/open-access-guidance/open-access-policy](https://wellcome.org/grant-funding/guidance/open-access-guidance/open-access-policy).] Accessed May 11, 2023.
- 384 8. We now accept preprints in grant applications: Wellcome; 2019 [Available from:
385 <https://wellcome.org/news/we-now-accept-preprints-grant-applications>.] Accessed May 11, 2023.
- 386 9. Chiarelli A, Johnson R, Pinfield S, Richens E. Preprints and Scholarly Communication: An
387 Exploratory Qualitative Study of Adoption, Practices, Drivers and Barriers [version 2; peer review: 3
388 approved, 1 approved with reservations]. *F1000Res.* 2019;8:971.
- 389 10. Kleinert S, Horton R. Preprints with The Lancet are here to stay. *Lancet.* 2020;396(10254):805.
- 390 11. Sheldon T. Preprints could promote confusion and distortion. *Nature.* 2018;559(7715):445.
- 391 12. van Schalkwyk F, Dudek J. Reporting preprints in the media during the COVID-19 pandemic.
392 *Public Underst Sci.* 2022;31(5):608-16.
- 393 13. Spungen H, Burton J, Schenkel S, Schriger DL. Completeness and Spin of medRxiv Preprint and
394 Associated Published Abstracts of COVID-19 Randomized Clinical Trials. *JAMA.* 2023;329(15):1310-2.
- 395 14. Bero L, Lawrence R, Leslie L, Chiu K, McDonald S, Page MJ, et al. Cross-sectional study of
396 preprints and final journal publications from COVID-19 studies: discrepancies in results reporting and
397 spin in interpretation. *BMJ Open.* 2021;11(7):e051821.
- 398 15. Brierley L, Nanni F, Polka JK, Dey G, Palfy M, Fraser N, et al. Tracking changes between preprint
399 posting and journal publication during a pandemic. *PLoS Biol.* 2022;20(2):e3001285.
- 400 16. Shi X, Ross JS, Amancharla N, Niforatos JD, Krumholz HM, Wallach JD. Assessment of
401 Concordance and Discordance Among Clinical Studies Posted as Preprints and Subsequently Published in
402 High-Impact Journals. *JAMA Netw Open.* 2021;4(3):e212110.

- 403 17. Itani D, Lababidi G, Itani R, El Ghoul T, Hamade L, Hijazi ARA, et al. Reporting of funding and
404 conflicts of interest improved from preprints to peer-reviewed publications of biomedical research. *J Clin*
405 *Epidemiol.* 2022;149:146-53.
- 406 18. Otridge J, Ogden CL, Bernstein KT, Knuth M, Fishman J, Brooks JT. Publication and Impact of
407 Preprints Included in the First 100 Editions of the CDC COVID-19 Science Update: Content Analysis. *JMIR*
408 *Public Health Surveill.* 2022;8(7):e35276.
- 409 19. Zeraatkar D, Pitre T, Leung G, Cusano E, Agarwal A, Khalid F, et al. Consistency of covid-19 trial
410 preprints with published reports and impact for decision making: retrospective review. *BMJ Med.*
411 2022;1(1):e000309.
- 412 20. Anazco D, Nicolalde B, Espinosa I, Camacho J, Mushtaq M, Gimenez J, et al. Publication rate and
413 citation counts for preprints released during the COVID-19 pandemic: the good, the bad and the ugly.
414 *PeerJ.* 2021;9:e10927.
- 415 21. Abdill RJ, Blekhman R. Tracking the popularity and outcomes of all bioRxiv preprints. *eLife.*
416 2019;8:e45133.
- 417 22. Fraser N, Brierley L, Dey G, Polka JK, Pálffy M, Nanni F, et al. The evolving role of preprints in the
418 dissemination of COVID-19 research and their impact on the science communication landscape. *PLOS*
419 *Biology.* 2021;19(4):e3000959.
- 420 23. Akbaritabar A, Stephen D, Squazzoni F. A study of referencing changes in preprint-publication
421 pairs across multiple fields. *Journal of Informetrics.* 2022;16(2).
- 422 24. Use of preprint articles in prevention research – a mixed-methods approach: Center for Open
423 Science; 2021 [Available from: <https://osf.io/cnkdw/>.] Accessed May 11, 2023.
- 424 25. Use of preprint articles in prevention research – a mixed-methods approach - update: Center for
425 Open Science; 2022 [Available from: <https://osf.io/k9ur2/>.] Accessed May 11, 2023.
- 426 26. Reitz K. requests 2.29.0: Python Software Foundation; 2023 [Available from:
427 <https://pypi.org/project/requests/>.] Accessed May 11, 2023.
- 428 27. Richardson L. beautifulsoup4 4.12.2: Python Software Foundation; 2023 [Available from:
429 <https://pypi.org/project/beautifulsoup4/>.] Accessed May 11, 2023.
- 430 28. Solomon B. re101 0.4.0: Python Software Foundation; 2018 [Available from:
431 <https://pypi.org/project/re101/>.] Accessed May 11, 2023.
- 432 29. Prevention Research Defined: The National Institutes of Health (NIH); 2020 [Available from:
433 <https://prevention.nih.gov/about-odp/prevention-research-defined/>.] Accessed May 11, 2023.
- 434 30. Gartlehner G, Dobrescu A, Evans TS, Thaler K, Nussbaumer B, Sommer I, et al. Average effect
435 estimates remain similar as evidence evolves from single trials to high-quality bodies of evidence: a
436 meta-epidemiologic study. *J Clin Epidemiol.* 2016;69:16-22.
- 437 31. Silagy CA, Middleton P, Hopewell S. Publishing Protocols of Systematic Reviews Comparing What
438 Was Done to What Was Planned. *JAMA.* 2002;287(21):2831-4.
- 439 32. Journal Impact Factor List 2019 – JCR, Web Of Science (PDF, XLS): Journal Impact Factor; 2020
440 [Available from: <https://impactfactorforjournal.com/journal-impact-factor-list-2019/>.] Accessed May 11,
441 2023.

- 442 33. Lu H, Cole SR, Platt RW, Schisterman EF. Revisiting Overadjustment Bias. *Epidemiology*.
443 2021;32(5).
- 444 34. Cole SR, Platt RW, Schisterman EF, Chu H, Westreich D, Richardson D, et al. Illustrating bias due
445 to conditioning on a collider. *International Journal of Epidemiology*. 2009;39(2):417-20.
- 446 35. Mansournia MA, Nazemipour M, Etminan M. Interaction Contrasts and Collider Bias. *American*
447 *Journal of Epidemiology*. 2022;191(10):1813-9.
- 448 36. Bürkner P-C. brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical*
449 *Software*. 2017;80(1):1 - 28.
- 450 37. Wickham H, Averick M, Bryan J, Chang W, McGowan L, François R, et al. Welcome to the
451 Tidyverse. *Journal of Open Source Software*. 2019;4(43):1686.
- 452 38. Wickham H, Bryan J. readxl: Read Excel Files: Comprehensive R Archive Network (CRAN); 2023
453 [Available from: <https://CRAN.R-project.org/package=readxl>.] Accessed May 11, 2023.
- 454 39. Kay M. tidybayes: Tidy Data and Geoms for Bayesian Models (v3.0.3): Zenodo; 2023 [Available
455 from: <https://doi.org/10.5281/zenodo.7606324>.] Accessed May 11, 2023.
- 456 40. Kay M. tidybayes: Tidy Data and Geoms for Bayesian Models: Zenodo; 2023 [Available from:
457 <http://mjskay.github.io/tidybayes/>.] Accessed May 11, 2023.
- 458 41. medRxiv: Cold Spring Harbor Laboratory (CSHL); 2023 [Available from:
459 <https://www.medrxiv.org/>.] Accessed May 11, 2023.
- 460 42. Else H. How a torrent of COVID science changed research publishing - in seven charts. *Nature*.
461 2020;588(7839):553.
- 462 43. Kodvanj I, Homolak J, Virag D, Trkulja V. Publishing of COVID-19 preprints in peer-reviewed
463 journals, preprinting trends, public discussion and quality issues. *Scientometrics*. 2022;127(3):1339-52.
- 464 44. Oikonomidi T, Boutron I, Pierre O, Cabanac G, Ravaud P, Consortium C-N. Changes in evidence
465 for studies assessing interventions for COVID-19 reported in preprints: meta-research study. *BMC Med*.
466 2020;18(1):402.
- 467 45. Carneiro CFD, Queiroz VGS, Moulin TC, Carvalho CAM, Haas CB, Rayee D, et al. Comparing quality
468 of reporting between preprints and peer-reviewed articles in the biomedical literature. *Res Integr Peer*
469 *Rev*. 2020;5(1):16.
- 470 46. Garcia-Costa D, Squazzoni F, Mehmani B, Grimaldo F. Measuring the developmental function of
471 peer review: a multi-dimensional, cross-disciplinary analysis of peer review reports from 740 academic
472 journals. *PeerJ*. 2022;10:e13539.
- 473 47. Sadler K. The Covid-19 outbreak highlights the potential of preprints: *Times Higher Education*;
474 2020 [Available from: [https://www.timeshighereducation.com/opinion/covid-19-outbreak-highlights-](https://www.timeshighereducation.com/opinion/covid-19-outbreak-highlights-potential-preprints)
475 [potential-preprints](https://www.timeshighereducation.com/opinion/covid-19-outbreak-highlights-potential-preprints).] Accessed May 11, 2023.
- 476