

1 Original Investigation












2

3 **Questionable practices in data and code sharing policy in high-profile medical journal and**
4 **research**

5

6 **Running title:** Failures of data/code sharing policy in medical journals

7

8 Wei Li, MBBS^{1,2†} , Xuerong Liu, Msc^{2†} , Qianyu Zhang, Msc² , Liping Shi, Msc², Jing-Xuan
9 Zhang, MD² , Xiaolin Zhang, Msc², Jia Luan, Msc³ , Yue Li, Msc³ , Ting Xu, PhD⁴ , Rong
10 Zhang, PhD⁴, Xiaodi Han, Msc², Jingyu Lei, Msc², Xueqian Wang, Msc², Yaozhi Wang, Msc⁵ , Hai
11 Lan, PhD⁶, Xiaohan Chen, PhD⁷, Yi Wu, Msc⁸, Yan Wu, BS⁹, Lei Xia, Msc², Haiping Liao, PhD²,
12 Chang Shen, Msc², Yang Yu, Msc², Xinyu Xu, Msc², Chao Deng, Msc², Pei Liu, Msc², Zhengzhi Feng,
13 MD² , Chun-Ji Huang, MD¹⁰ , Zhiyi Chen, PhD^{1,2,4,11*} 

14

15 **Affiliation:**

16 ¹ The Chinese Open Science Network, Nanjing, China 210023

17 ² Experimental Research Center for Medical and Psychological Sciences, Third Military Medical
18 University, Chongqing, China 400038

19 ³ Editorial Office, The Journal of Third Military Medical University, Chongqing, China 400038

20 ⁴ School of Psychology, Southwest University, Chongqing, China 400715

21 ⁵ School of Education, Sichuan Normal University, Chengdu China 610066

22 ⁶ School of Psychology, Sichuan Normal University, Chengdu, China 610066

23 ⁷ The Chengdu University of Traditional Chinese Medicine, Chengdu, China 611137

24 ⁸ School of Management, Third Military Medical University, Chongqing, China 400038

25 ⁹ School of Architecture , Zhengzhou University, Zhengzhou, China 450001

26 ¹⁰ President Office, Third Military Medical University, Chongqing, China 400038

27 ¹¹ MOE Key Laboratory of Cognition and Personality, Chongqing, China 400715

28

29 * Corresponding at Zhiyi Chen (PhD., PI, Senior Research Fellow; No. 30 Gao Tan-Yan Main Street,
30 Shapingba, Chongqing, China; TEL: +86 0 23 771767; Fax: +86 0 23 752341)

31 **Email:** chenzhiyi@tmmu.edu.cn

32

33 † These authors contributed equally to this work

34

35  **ORCID**

36 Wei Li, 0009-0004-1347-9302

37 Xuerong Liu, 0000-0002-9236-5773

38 Qianyu Zhang, 0009-0006-7606-2218

39 Jing-Xuan Zhang, 0000-0002-8979-5107

40 Jia Luan, 0000-0002-6258-5046

41 Yue Li, 0000-0002-7824-4246

42 Ting Xu, 0000-0001-9278-6474

43 Yaozhi Wang, 0009-0009-2889-0176

44 Zhengzhi Feng, 0000-0001-6144-5044

45 Chun-Ji Huang, 0000-0002-6108-539X

46 Zhiyi Chen, 0000-0003-1744-4647

47

48 **This file includes:**

49 2,959 Main Text (not including title, abstract, acknowledgment, references, tables, and figure legends)

50 3 of 3 Color figures

51 2 of 2 Tables

52 3 of 3 Appendix (Supplemental Information and raw data)

53 6 of 6 eTables in Supplement Online

54 **Abstract**

55 **Background**

56 The spurious and unavailable data/code sharing actions are crashing open medical sciences. In this
57 study, we aimed to illustrate how high-profile medical journals are practically carried out their sharing
58 policies and what questionable practices regarding data/code sharing are conducted by authors.

59

60 **Methods**

61 In this study, we appraised the policy on data/code availability of high-profile medical journals ranked
62 at Q1 according to Clarivate Journal Citation Report (JCR 2021). Furthermore, we recruited post-
63 publications published by four leading medical journals (i.e., The BMJ, JAMA, NEJM and The Lancet)
64 from the issuing of data/code availability policy to December 2022 for the questionable practices in
65 data/code sharing. The appraisal of papers was conducted by the Data/code Availability Statement
66 Practice Evaluation Tool (DANCE), developed by systematically integrating mainstreaming open
67 data/code guidelines.

68

69 **Findings**

70 We found that less than one-tenth journals (9.1%) mandated authors to share data/code, with an
71 available statement. Among these journals, 70.6% (61.2%) did not consider censoring (restricting)
72 spurious/invalid data/code sharing in publications. Furthermore, though journal impact factor could
73 predict policy stringency on “offering availability statements” ($p < .001$), it failed to predict ones in
74 “sharing data/code” ($p = .73$). For publications, even in leading medical journals (i.e., The BMJ, JAMA,
75 NEJM and The Lancet), only 0.5% of the papers (16/3,191) fully complied with their public sharing
76 statements for reaching reproducibility. Lack of availability statement, declining data/code sharing
77 without reasons, and invalid repositories were leading questionable practices conducted by authors.

78

79 **Interpretation**

80 We clarified specific questionable actions of implementing and practicing the sharing policy both in
81 journal and papers, which should be addressed not only by the supportive publication ecosystem but
82 also by crediting authors for taking responsibility and maintaining scientific integrity in data/code
83 sharing.

84

85 **Funding**

86 No funding.

87

88

89

90

91

92

93

94

95

96

97

98 **Introduction**

99 The importance of data and code sharing is evident, as it has been well-acknowledged to benefit the
100 best interests of science and public health good, particularly in promoting open, transparent,
101 reproducible and trustworthy sciences^{1, 2}. Despite such promising and ambitious goals, the actions
102 toward these grand promises are seriously challenged^{3, 4}. Though declaring to embrace and welcome
103 open data/code sharing initiatives/ethics, the prevalence of issuing data sharing policies in (even
104 leading) medical journals (i.e., < 10%) has not yet achieved open science goals, with no prominent
105 changes resulting from the implementation of institutional or stakeholder-driven data sharing
106 statements in the last decade⁵. Moreover, in these journals with policies requiring or even mandating
107 sharing, 98.0% (99.5%) of papers do not actually share available research data (code), showing a
108 notorious gap between declared policies and actual publication practices⁶. However, for these
109 challenges, we know very little about how these journal policies are actually carried out, and what
110 questionable practices are exactly conducted by authors in these publications.

111
112 Journals are the first gatekeeper to prevent spurious or unavailable data/code sharing practices. Many
113 scientific communities such as The Transparency and Openness Promotion (TOP)⁷ and the
114 Recommendation on Open Science from United Nations Educational, Scientific and Cultural
115 Organization (EOP-UNESCO)⁸ offer structural guidelines to formulate open data/code policies at the
116 journal side, clearly indicating step-by-step censorship to data/code sharing integrity with certified
117 standards (e.g., FAIR data stewardship)⁹. However, neither supportive policy changes nor increased
118 priority to data sharing practices have occurred after proclaiming these guidelines. Beyond such
119 community-based recommended guidelines, the International Committee of Medical Journal Editors
120 (ICMJE) has ever imposed to report data sharing statement/plan to clinical trial data¹⁰, but nothing has
121 changed in the prevalence of data sharing¹¹. Despite the consensus on poor prevalence, it has still been
122 underexplored regarding how journals practically carry out their sharing policies. For example, *did*
123 *journals merely require a statement to report data sharing methods, but not require actual data sharing?*
124 *or did high-rank journals guarantee stringent policy on data/code sharing?*

125
126 For the policy-to-practice gap, though spurious and actually unavailable data/code sharing statements
127 are incredibly pervasive^{6, 12}, we can do nothing to address this notorious deterioration until we
128 understand what (intentionally or inadvertently) questionable practices are conducted by authors¹³. On
129 the one hand, clarifying these specific questionable research actions may contribute to guiding journals
130 in tailoring sharing policies, with add-on clauses, to self-correct policy loopholes for actually valid and
131 available sharing¹⁴. Moreover, such findings may additionally benefit editorial or peer review in
132 censoring data/code sharing integrity, indicating notable risky points in their data/code sharing
133 statements¹⁵. On the other hand, for authors, identifying reasons of incurring failure on actually
134 available data/code sharing in their publication practices could further educate authors on how to adjust
135 sharing strategies when the journal policy is not yet supportive/helpful enough. Supporting this
136 argument, the poor data/code sharing implementation has been partly attributed to the low practicability
137 in these guidelines/policies per se¹⁶.

138
139 Here, to answer the first question, we utilized the data/code availability policy (DAP) matrix (**Table 1**)
140 to appraise degree of policy on data/code availability in 931 high-profile medical journals ranked at Q1
141 according to Clarivate Journal Citation Report (JCR 2021). To answer the second question, for the four

142 leading medical journals (i.e., British Medical Journal, BMJ; Journal of the American Medical
143 Association, JAMA; New England Journal of Medicine, NEJM; Lancet), we evaluated 3,191 papers
144 published from the issuing of data/code availability policy to December 2022 for the questionable
145 practices in data/code sharing/availability. This meta-research appraisal was conducted by the
146 Data/code Availability Statement Practice Evaluation Tool (DANCE) , which was developed by
147 systematically integrating mainstreaming open data/code guidelines (**eTables 2-6**).

148 **Methods**

149 **Study design**

151 The current study employed meta-research to appraise the implementation of high-profile medical
152 journal policies on data/code sharing and availability statements. Furthermore, the study also aimed to
153 assess questionable data/code sharing risks and to identify specific questionable practices impeding
154 data/code sharing in papers published in leading medical journals.

155 **Search strategy and eligibility criteria**

157 For journal policies, we began by selecting all medical journals ranked at Quartile 1 (Q1) in the
158 category of Clinical Medicine from Clarivate Journal Citation Report (JCR) 2021 on 1st June 2023.
159 Journals publishing original research were included in the present study, while review journals and
160 book series were excluded. We finally included 931 journals for the appraisal of policies.

162 For data/code availability practices in papers, we first included all papers published in the leading
163 journals including the British Medical Journal (The BMJ), Journal of the American Medical
164 Association (JAMA), New England Journal of Medicine (NEJM) and The Lancet, restricting the time
165 from issuing sharing policy (2018 for BMJ and JAMA; 2019 for NEJM; 2021 for The Lancet) to 2022.
166 All of the leading journals required authors to provide an availability statement to disclose whether the
167 data/code would be shared for specific paper types. All specific article types which were required to
168 provide such statement by journals were included, such as original articles and brief reports, while case
169 report, review, abstract and qualitative research were excluded. Finally, 3,191 papers were included in
170 the current analysis.

172 Two reviewers (L.W. and L.X.R.) independently checked the eligibility for the included journals and
173 papers by screening the journals' aim, scope and guidelines for authors, as well as all papers' types,
174 titles, abstracts and main texts. Any disagreements were resolved by a senior reviewer (C.Z.Y.).

175 **Data collection and processing**

177 For eligible journals, the journal characteristics, including Uniform Resource Locator (URL), 2021
178 impact factor, publisher, subspecialty, publication frequency, International Standard Serial Number
179 (ISSN), and electronic ISSN, were firstly extracted from Clarivate Journal Citation Report (JCR 2021)
180 by two reviewers (L.W. and L.X.R.). For journals belonging to multiple subspecialties, we counted
181 them in each subspecialty, in the subgroup analysis of subspecialties. Then, two reviewers (Z.Q.Y. and
182 W.Y.Z.) systematically extracted the descriptions regarding the data/code sharing and availability
183 statement recorded in Guidelines for Authors manually for appraising the implementation of journal
184 policies.

185

186 For eligible papers, two reviewers (H.X.D. and L.J.Y.) manually extracted bibliographic characteristics
187 (journal, title, published year, article type, DOI, study design) and data/code availability statement for
188 assessing questionable data/code sharing risks and to identify specific questionable practices impeding
189 data/code sharing as guided by the DANCE tool (see below “**Data Analysis**”).

190

191 **Data Analysis**

192 In the analysis of journals’ policies, we built on a data/code availability policy (DAP) matrix (**Table 1**).
193 The degree of each journal was classified into four ranks based on the policy description regarding
194 data/code sharing and availability statement. “No policy” denotes the complete absence of any
195 directives or guidelines regarding data/code sharing and availability statements. “Encouraged” denotes
196 that the decision to share data/code or provide an availability statement is left to the discretion of the
197 authors. “Required” denotes that the journal requires authors to include an availability statement
198 specifying whether data/code will be shared. “Mandated” denotes that the journal obliges authors to
199 both share data/code and provide an availability statement detailing the methods through which the
200 shared data/code can be accessed by others. The degree of each journal depends on the description of
201 policies. We further conducted subgroup analysis of journals’ policies, including subspecialties,
202 publishers and journal impact factors. Three reviewers (X.X.Y., D.C. and L.P.) independently rated the
203 policies, with disagreement solved by the fourth author (C.Z.Y.).

204

205 In the analysis of papers’ practices, the Data/code Availability Statement Practice Evaluation Tool
206 (DANCE) was developed by systematically to assessing integrating and structuring mainstream open
207 data/code guidelines, which was developed to assess questionable data/code sharing risks and to
208 identify specific questionable practices impeding data/code sharing, from four domains based on the
209 process of reproducing the results of papers (**eTables 2-6**). The first domain (statement integrity) refers
210 to the comprehensiveness of the statement, ensuring it encompasses all key elements as stipulated by
211 the journal’s policy. The second domain (actual accessibility) refers to the practical implementation of
212 the availability statement, confirming that the data/code are indeed accessible as originally claimed.
213 The third domain (user usability) refers to the retrieved data/code are organized in a manner that
214 facilitates user-friendly verification. The fourth domain (method practicability) refers to whether the
215 original results can be fully reproduced utilizing the accessed data/code. In each domain, reviewers are
216 required to answer several signaling questions with “yes”, “probably yes”, “no”, “probably no” or
217 “unclear/not applicable”. Each domain was rated as high risk of questionable data/code sharing
218 practices if 1 or more items were answered with no/probably no. Only applicable articles could be rated
219 in each domain. For example, when a paper declares not to share its data/code, it is not applicable to
220 Domains 2,3, and 4, and would not be rated in these domains. Nine authors (L.W., L.X.R, Z.Q.Y.,
221 L.J.Y., H.X.D., W.Y.Z., W.X.Q., S.C. and Y.Y.) independently rated the policies, with disagreement
222 solved by the tenth author (C.Z.Y.).

223

224 **Statistical analysis**

225 We utilized descriptive statistical analyses in Microsoft Excel 2021, presenting these data as
226 frequencies and rates. Furthermore, in the subgroup analysis of journal impact factor, we used
227 Spearman correlation analysis conducted by SPSS (IBM, Inc., version 29.0.1.0).

228

229 **Results**

230 **Journal policies on data/code sharing**

231 Of 931 journals, 42.2% (393/931) required a statement disclosing data/code availability, 28.7%
232 (267/931) did not require it (i.e., encouraged/optional), and 29.1% (271/931) did not include any
233 clauses (**Fig. 1 a**). For data/code sharing policy, more than half of the journals (601/931, 64.6%) did not
234 require authors to share data/code (i.e., encouraged/optional), 22.6% (210/931) did not include any
235 clauses, and only 12.9% (120/931) required data/code sharing (**Fig. 1 b**). In summary, in these high-
236 profile medical journals claiming embrace of open data initiatives/ethics, less than one-tenth of the
237 journals (85/931, 9.1%) mandated authors to share research data/code along with a clear availability
238 statement (**Fig. 1 c**). To make matters worse, even among the 85 journals with mandatory data
239 availability policies, 60 journals (70.6%) had not yet disclosed any implemented measures in their
240 policies to ensure that data/code shared in publications were genuine and valid, and 52 journals (61.2%)
241 had not considered making sharing of spurious or invalid data/code a disqualifying condition for
242 publication.

243

244 In subgroup analyses, by categorizing into subspecialties and publishers, we further examined the
245 above proportions of journals' policies across 59 medical subspecialties and 13 publishers (containing
246 at least 10 journals). Regarding the subspecialties, subspecialties with over 20% of journals
247 implementing mandatory data/code sharing and statement policies accounted for 18.6% (11/59) (**Fig. 1**
248 **d**). As for the publishers (**Fig. 2 a**), the proportion of such mandatory policies varied largely from 0% to
249 50%. Finally, we conducted Spearman correlation analysis to investigate whether the 2021 Journal
250 Impact Factors (JIFs) could predict the policy compliance. Interestingly, though the journal ranks
251 quantified by 2021 JIFs could predict policy stringency on "offering data availability statement"
252 ($\rho=0.20$, 95% CI: 0.13-0.26, $p<0.001$), it failed in predicting the policies on "sharing actual data/code"
253 ($\rho=0.01$, 95% CI: -0.06-0.08, $p=0.737$; **Fig. 2 b**), which possibly implied the risk of the practice of
254 formalities for formalities' sake in promoting research transparency.

255

256 **The risk of conducting questionable data/code sharing**

257 When examining data/code sharing statements for individual 3,191 papers, 92.3% (1,555/1,685) of
258 those with private availability and 95.8% (413/431) with public availability were rated as high risk for
259 questionable sharing practices. The remaining 33.7% (1,075/3,191) papers lacked the statements. For
260 each specific appraisal domain in the DANCE (**Table 2**), 93.3% were rated as high risk for statement
261 integrity, 46.1% for actual accessibility, 89.5% for user usability, and 76.5% for method practicality.
262 Consequently, only 0.5% (16/3,191) could be empirically reproduced by following their sharing
263 statements.

264

265 **Specific questionable practices impeding data/code sharing**

266 Given the pervasive risk of questionable data/code sharing identified above, We synthesized
267 questionable practices from "high-risk" papers and found that one-third of the papers (1,075 of 3,191,
268 33.7%) did not include any descriptions about data/code availability, despite the clear requirements of
269 availability statements by journals (**Fig. 3**). Though providing such statements, 23.3% (745/3,191)
270 refused to share data/code, in which 58.4% (435/745) did not provide any reasons. Even worse, 13.5%
271 (431/3,191) declared to share data/code publicly, but 39.0% (168/431) were not reachable to the
272 sharing data/code. The remaining 29.5% (940/3,191) papers declared to share data/code upon request,
273 with neither clarification to permission conditions nor explanations to the data/code preservation.

274

275 **Discussion**

276 Data and code sharing policy certainly fueled open, transparent and better science, but not as much as
277 we expected. In this meta-research study, we found the journal policies and grandiloquent promises did
278 not guarantee scientific integrity on open science practices. Even in the high-profile medical journals,
279 less than one tenth of them declared a mandate for authors to share data/code, along with a clear
280 availability statement. Moreover, these journals neither stipulated censorship of these statements
281 (70.6%) nor enforced valid data/code sharing as an obligatory condition of publication (61.2%). Even
282 worse, such questionable practices in the journal policy on data/code sharing seemed to be pervasive,
283 as we observed no significant associations between journal ranks (i.e., journal impact factor) and policy
284 stringency. Compared to the motivation for declaring a data/code sharing policy, actual efforts to
285 mandate or impose this practice on authors have been less prioritized. Beyond journal policy flaws, we
286 found that authors (intentionally or inadvertently) conducted questionable practices on impeding valid
287 data/code sharing, in leading medical journals, particularly in lacking sharing statements, declining
288 data/code sharing without reasonable explanations, and invalidating data/code repositories. Despite the
289 acknowledged low actual data/code sharing quality and low policy compliance^{6, 12}, we exactly clarified
290 what practices are conducted to hamper readers actually accessing their data/codes, which were
291 suggested to be addressable at journal sectors or at authors per se.

292

293 For journal policy, though scientific communities and publishers vociferously promoted open science
294 by supporting “data/code sharing policy”, our findings suggested that, most high-profile medical
295 journals have merely taken “politically-correct” actions, with a formalistic (but not actual)
296 requirements on policy practices. As previously reported, the editors and leading medical publishers
297 (i.e., The Lancet) did not value or prioritize the quality of data/code sharing, as they have no editorial
298 responsibilities for research practices but possess duties on publication ethics^{17, 18}. Thus, such
299 “meretricious” policies may readily become an “empty promise” to readers, for being posturing to
300 endorse open sciences only⁴. Furthermore, though leading medical journals (e.g., The BMJ) indeed
301 declared to take responsibility for data/code sharing¹⁹, they contributed not much to improve its validity,
302 with less than 10% of these papers fully sharing research data/code. Therefore, it is evident that taking
303 editorial responsibilities to implement mandatory policies alone is not enough to address questionable
304 data/code sharing problems²⁰. Instead, a supportive publication ecosystem including peer review,
305 scientific integrity education, institutional funds, and technical censorship to data/code sharing
306 practices is more imperative²¹⁻²³.

307

308 For the overarching stakeholder - authors, despite their ostensible enthusiasm for open science, one
309 disappointing truth is that many failed to responsibly share their data/code enough for reaching
310 transparency and reproducibility. To our knowledge, beyond well-documented studies on the
311 motivations impeding authors from sharing data/code (e.g., technical errors, high time costs, less
312 credits/benefits, or competitive disadvantage)^{24, 25}, this is the first study to clarify how authors may
313 questionably practice data/code sharing policy in pursuing publication. One primary questionable
314 conduct is not offering data/code sharing statements or restricting access upon request without
315 reasonable considerations. This suggests that, rather than putting all trust on proactive actions from
316 authors, compliance with data/code sharing policy should be prescribed as authors’ responsibility,
317 which is a reflection of scientific integrity and even publication ethics per se^{13, 26}. By doing so, the

318 data/code sharing policy no longer earns open science credits for journals but directly benefits authors’
319 research integrity²⁷. Furthermore, credit or incentive to such actions is the key²⁸. As indicated by the
320 TOP guideline, the “open data” badges, scientific integrity credits or other awards should be given to
321 acknowledge that authors are taking the responsibility²⁹. Another main questionable practice is to
322 (intentionally or inadvertently) invalidate data/code sharing repositories while declaring public access.
323 Technically speaking, this flaw could be readily addressed by technical scrutiny. For instance, an end-
324 to-end text-mining model has been applied to automatically detect data sharing integrity, such as
325 scrutinizing whether research data have been clearly claimed, fully accessed, actually downloaded and
326 practically operated³⁰. At authors’ side, crediting their responsibility and scientific integrity on
327 data/code sharing, along with stringent technical checks, may be a practical solution.

328
329 Here, several limitations should be acknowledged. First, we narrowed the scope of probing data/code
330 sharing policies to these high-profile medical journals, which alluded to the hypothesis that these were
331 representative of stricter and high-quality policies in the medically science than others. Despite this
332 premise has been substantially supported in previous evidences, a landscape of policy and its practices
333 in all the medical journals (> 7000) could strengthen the reliability and generalizability of these
334 findings. Second, we focused on the practical quality of public data/code sharing, and did not send data
335 requests to validate practices in the statements declaring “upon request”. Third, given the lack of
336 standardized and structural tool to systematically appraise practicability of data/code sharing, we
337 tailored the DANCE by systematically integrating and structuring these mainstream open data/code
338 guidelines. These findings could be further validated once a reliable checklist/tool was prepared.

339
340 In conclusion, journals’ policies to support transparent, open and better science are unprecedentedly
341 ambitious and undoubtedly conducive to benefiting the medical community, but the actual
342 effectiveness remains very suboptimal. Despite issuing data/code sharing policy, journals/publishers
343 may merely focus on “declaring” rather “sharing” research data/code in the publication practices,
344 without censorship and restriction to spurious/substandard “data availability” statements. Even in
345 leading journals, under data sharing requirements in the policies, the questionable data/code sharing
346 practices impeding data access in the papers are substantially pervasive. Authors may intentionally (or
347 inadvertently) conduct multifarious questionable practices to repudiate data/code sharing. Such
348 potential misconducts in practicing data/code sharing should have been addressed not only by
349 establishing a supportive publication ecosystem but also by crediting authors for taking responsibility
350 and maintaining scientific integrity in data/code sharing.

351 **Acknowledgments**

352 we do thank Daniel G Hamilton for supports on data curation.

353

354 **Author Contributions**

355 Zhiyi Chen, Xuerong Liu and Wei Li: Conceptualization, Methodology and Writing - Original Draft;
356 Chun-Ji Huang, Jia Luan, Yue Li, and ZhengzhiFeng: Conceptualization; Liping Shi, Jing-Xuan Zhang,
357 Ting Xu, Rong Zhang and Xiaolin Zhang: Writing - Review & Editing; Wei Li, Xuerong Liu, Qianyu
358 Zhang, Xiaodi Han, Jingyu Lei, Xueqian Wang, Yaozhi Wang, Hai Lan, Xiaohan Chen, Yi Wu, Yan Wu,
359 Lei Xia, Haiping Liao, Chang Shen, Yang Yu, Xinyu Xu, Chao Deng and Pei Liu: Data Curation &
360 Validation; Zhiyi Chen: Project Administration and Funding Acquisition. L.W. and L.X.R. shared first
361 authorship.

362

363 **Competing Interest Statement**

364 All the authors disclosed no conflicts of interests.

365

366 **Data availability**

367 The datasets of journal policy and article practice and the code required to reproduce all of the findings
368 of this meta-research are available on the Open Science Framework (OSF) (<https://osf.io/5m3zc/>).

369 **Tables and Figures**

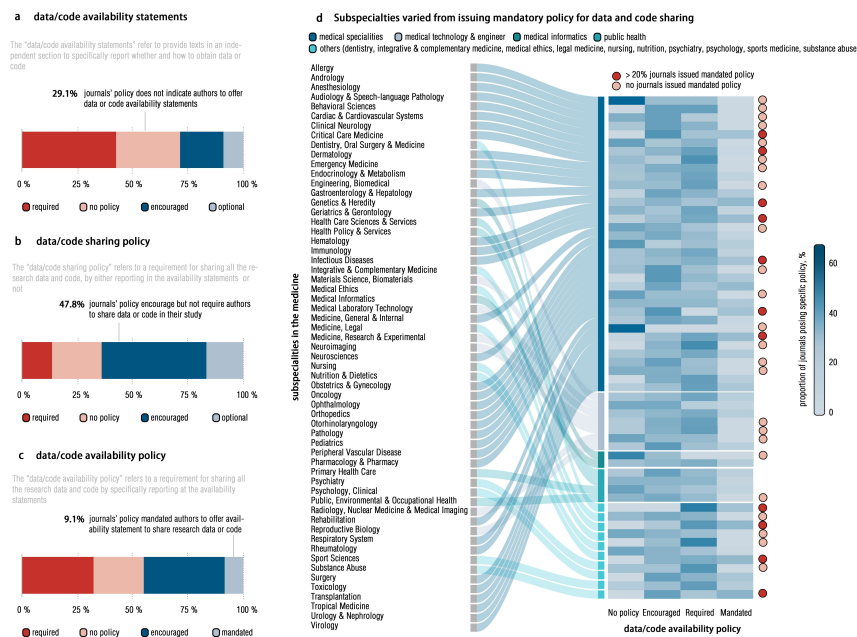
370

	Data/code Availability Statement Policy		
	No policy	Optional/Encouraged	Required
Data/code Sharing Policy			
No policy	No policy	Encouraged	Required
Optional/Encouraged	Encouraged	Encouraged	Required
Required	Encouraged	Encouraged	Mandated

371 **Table 1: The Data/code Availability Policy (DAP) matrix.** The degree of data/code availability policy of journal primarily depends on the requirements regarding data
372 sharing policy and data availability statement policy. Note: both “optional” and “encouraged” in data/code sharing policy and availability policy refer to that the willingness
373 of sharing data or providing availability statement depends on the authors themselves, but differing in the strength of the suggestion, with “encouraged” being more strongly
374 worded than “optional”.

	Yes/probably yes	No/probably no	Unclear/Not Applicable
Domain 1: Statement Integrity	213 (6.7%) ^a	2,978 (93.3%) ^b	0
1.1 Was this paper provided Data Availability Statement or/and Code (Scripts) Availability Statement or/and Materials Availability Statement.	2,124 (66.6%)	1,067 (33.4%)	0
1.2 Was this paper explained why data, code (scripts) or materials sharing were under restrictions when it claimed “upon reasonable request” or “no additional data available”?	952 (29.8%)	567 (17.8%)	1,672 (52.4%)
1.3 Were these data, code (scripts) or materials deposited in a reliable repository, with specific hyperlinks or entrances?	179 (5.6%)	170 (5.3%)	2,842 (89.1%)
1.4 Was this repository has unique and persistent identifiers (e.g., DOI)?	181 (5.7%)	1 (0.0%)	3,009 (94.3%)
1.5 Was this Data Availability Statement or/and Code (Scripts) Availability Statement or/and Materials Availability Statement discoverable?	2,124 (66.6%)	0	1,067 (33.4%)
1.6 Were the data, code (scripts) or materials sharing integral?	248 (7.8%)	1,857 (58.2%)	1,086(34.0%)
Domain 2 Actual Accessibility	104 (53.9%) ^a	89 (46.1%) ^b	2,998
2.1 Were these hyperlinks or entrances valid actually?	176 (5.5%)	17 (0.5%)	2,998 (94.0%)
2.2 Were the permissions or accesses actually given for all the users?	112 (3.5%)	64 (2.0%)	3,015 (94.5%)
2.3 Were these accessing data, code (scripts) or materials actually conformed to statement?	109 (3.4%)	8 (0.3%)	3,074 (96.3%)
2.4 Were these accessing data, code (scripts) or materials actually obtained?	115 (3.6%)	0	3,076 (96.4%)
2.5 Were these obtained files workable?	114 (3.6%)	1 (0.0%)	3,077 (96.4%)
Domain 3: User Usability	12 (10.5%) ^a	102 (89.5%) ^b	3,077
3.1 Were these accessed data/code/materials indicated for how to use?	42 (1.3%)	72 (2.3%)	3,077 (96.4%)
3.2 Were these accessed data/code/materials organized user-friendly?	108 (3.4%)	6 (0.2%)	3,077 (96.4%)
3.3 Were the software or statistical toolkit prepared to use these data/code/materials?	62 (1.9%)	52 (1.6%)	3,077 (96.4%)
3.4 Were these data/code/materials understandable for all the users (e.g., junior, trained and senior researchers)?	13 (0.4%)	101 (3.2%)	3,077 (96.4%)
Domain 4: Method Practicability	16 (23.5%) ^a	52 (76.5%) ^b	3,123
4.1 Were the programming or technical environments generalizable for users?	67 (2.1%)	1 (0.0%)	3,123 (97.9%)
4.2 Were these codes/scripts actually workable?	12 (0.4%)	12 (0.4%)	3,167 (99.2%)
4.3 Were the necessary comments or information were provided within code or scripts?	54 (1.7%)	4 (0.1%)	3,133 (98.2%)
4.4 Were the dependent software/packages/toolkit provided?	59 (1.8%)	8 (0.3%)	3,124 (97.9%)
4.5 Were these code/scripts editable?	11 (0.3%)	14 (0.4%)	3,166 (99.2%)
Reproducibility	16 (0.5%) ^c		

376 **Table 2: The risk of questionable data/code sharing practices for papers using Data/code Availability Statement Practice Evaluation Tool (DANCE) (N=3,191).** Only
377 applicable articles could be rated in each domain. For example, when a paper declares not to share its data/code, it is not applicable to Domains 2,3, and 4, and would not be
378 rated in these domains. Each domain was rated as high risk of questionable data/code sharing practices if 1 or more items were answered with no/probably no. a. low risk of
379 questionable data/code sharing practices. b. high risk of questionable data/code sharing practices. Articles that were not applicable were excluded from the risk proportion
380 calculations. c. fully reproduced papers.



381

382

Fig. 1: The proportions of policy degrees for all Q1 journals regarding data/code availability. (a) shows

383

the degree of the requirement for data/code availability statement in publications. (b) shows the degree

384

of the requirement for data/code sharing in publications. (c) shows the overall degree concerning both

385

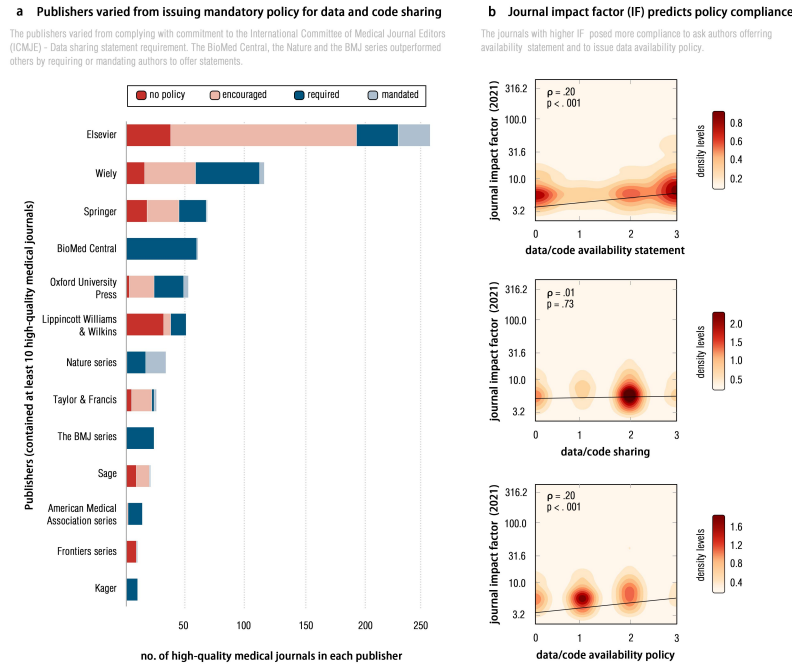
availability statement and data/code sharing. (d) illustrated the distribution of overall degree across

386

subspecialties (n=59) in medical science. For journals belonging to multiple subspecialties, we counted

387

them in each subspecialty.



388

389

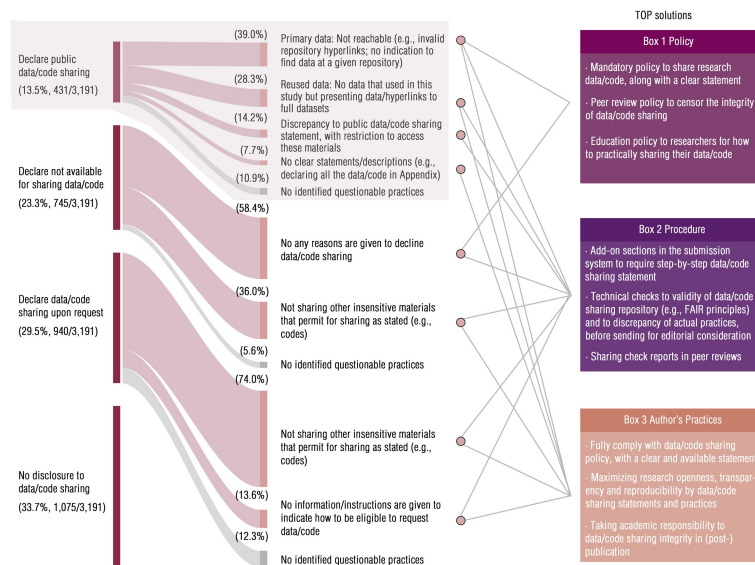
390

391

392

393

Fig. 2: The subgroup analyses in publishers and journal impact factors. (a) shows the distribution of overall degree across publishers. We only included publishers containing at least 10 high-profile medical journals. (b) shows the correlation between the policy of data/code availability statement, data/code sharing or overall degree (data/code availability) and journal impact factor (2021) according to the Clarivate Journal Citation Report (JCR) 2021, estimated by Spearman Correlation Analysis.



394

395 **Fig. 3:** The questionable data/code availability practices in publications and solutions according to the
 396 Transparency and Openness Promotion (TOP) Guideline. The practices were structured based on the
 397 declared data/code availability and actual data/code availability within each type of declared
 398 availability. The TOP solutions were divided into journal policy, submission procedure, and author's
 399 practices as outlined in the TOP Guideline.

400 **References**

- 401 1. Abbasi K. A commitment to act on data sharing. *BMJ (Clinical research ed)*. 2023;**382**:p1609.
402 10.1136/bmj.p1609.
- 403 2. Doshi P, Godlee F, Abbasi K. Covid-19 vaccines and treatments: we must have raw data, now.
404 *BMJ (Clinical research ed)*. 2022;**376**:o102. 10.1136/bmj.o102.
- 405 3. Page MJ, Nguyen PY, Hamilton DG, Haddaway NR, Kanukula R, Moher D, et al. Data and code
406 availability statements in systematic reviews of interventions were often missing or inaccurate: a
407 content analysis. *J Clin Epidemiol*. 2022;**147**:1-10. 10.1016/j.jclinepi.2022.03.003.
- 408 4. Gabelica M, Bojčić R, Puljak L. Many researchers were not compliant with their published data
409 sharing statement: a mixed-methods study. *J Clin Epidemiol*. 2022;**150**:33-41.
410 10.1016/j.jclinepi.2022.05.019.
- 411 5. Stodden V, Seiler J, Ma Z. An empirical analysis of journal policy effectiveness for computational
412 reproducibility. *PNAS*. 2018;**115**(11):2584-9. doi:10.1073/pnas.1708290115.
- 413 6. Hamilton DG, Hong K, Fraser H, Rowhani-Farid A, Fidler F, Page MJ. Prevalence and predictors
414 of data and code sharing in the medical and health sciences: systematic review with meta-analysis of
415 individual participant data. *BMJ (Clinical research ed)*. 2023;**382**:e075767. 10.1136/bmj-2023-075767.
- 416 7. Nosek BA, Alter G, Banks GC, Borsboom D, Bowman SD, Breckler SJ, et al. Promoting an open
417 research culture. *Science*. 2015;**348**(6242):1422-5. doi:10.1126/science.aab2374.
- 418 8. UNESCO. UNESCO Recommendation on Open Science. Paris: UNESCO; 2021.
- 419 9. Wilkinson MD, Dumontier M, Aalbersberg IJ, Appleton G, Axton M, Baak A, et al. The FAIR
420 Guiding Principles for scientific data management and stewardship. *Sci Data*. 2016;**3**(1):160018.
421 10.1038/sdata.2016.18.
- 422 10. Taichman DB, Sahni P, Pinborg A, Peiperl L, Laine C, James A, et al. Data Sharing Statements for
423 Clinical Trials: A Requirement of the International Committee of Medical Journal Editors. *Jama*.
424 2017;**317**(24):2491-2. 10.1001/jama.2017.6514.
- 425 11. Bergeat D, Lombard N, Gasmi A, Le Floch B, Naudet F. Data Sharing and Reanalyses Among
426 Randomized Clinical Trials Published in Surgical Journals Before and After Adoption of a Data
427 Availability and Reproducibility Policy. *JAMA network open*. 2022;**5**(6):e2215209.
428 10.1001/jamanetworkopen.2022.15209.
- 429 12. Naudet F, Sakarovitch C, Janiaud P, Cristea I, Fanelli D, Moher D, et al. Data sharing and
430 reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy:
431 survey of studies published in The BMJ and PLOS Medicine. *BMJ (Clinical research ed)*.
432 2018;**360**:k400. 10.1136/bmj.k400.
- 433 13. Federer LM, Lu Y-L, Joubert DJ, Welsh J, Brandys B. Biomedical Data Sharing and Reuse:
434 Attitudes and Practices of Clinical and Scientific Research Staff. *PloS one*. 2015;**10**(6):e0129506.
435 10.1371/journal.pone.0129506.
- 436 14. Hardwicke TE, Mathur MB, MacDonald K, Nilsonne G, Banks GC, Kidwell MC, et al. Data
437 availability, reusability, and analytic reproducibility: evaluating the impact of a mandatory open data
438 policy at the journal Cognition. *Royal Society open science*. 2018;**5**(8):180448. 10.1098/rsos.180448.
- 439 15. Munafò MR, Nosek BA, Bishop DVM, Button KS, Chambers CD, Percie du Sert N, et al. A
440 manifesto for reproducible science. *Nature Human Behaviour*. 2017;**1**(1):0021. 10.1038/s41562-016-
441 0021.
- 442 16. Kianersi S, Grant SP, Naaman K, Henschel B, Mellor D, Apte S, et al. Evaluating Implementation
443 of the Transparency and Openness Promotion Guidelines: Reliability of Instruments to Assess Journal

- 444 Policies, Procedures, and Practices. *Adv Methods Pract Psychol Sci*. 2023;**6**(1):25152459221149735.
445 10.1177/25152459221149735.
- 446 17. Naaman K, Grant S, Kianersi S, Supplee L, Henschel B, Mayo-Wilson E. Exploring enablers and
447 barriers to implementing the Transparency and Openness Promotion Guidelines: a theory-based survey
448 of journal editors. *Royal Society open science*. 2023;**10**(2):221093. 10.1098/rsos.221093.
- 449 18. Horton R. Offline: Data sharing-why editors may have got it wrong. *Lancet* (London, England).
450 2016;**388**(10050):1143. 10.1016/s0140-6736(16)31638-5.
- 451 19. Loder E, Macdonald H, Bloom T, Abbasi K. Mandatory data and code sharing for research
452 published by The BMJ. *BMJ* (Clinical research ed). 2024;**384**:q324. 10.1136/bmj.q324.
- 453 20. Tedersoo L, Küngas R, Oras E, Köster K, Eenmaa H, Leijen Ä, et al. Data sharing practices and
454 data availability upon request differ across scientific disciplines. *Scientific Data*. 2021;**8**(1):192.
455 10.1038/s41597-021-00981-0.
- 456 21. Milkman KL, Berger J. The science of sharing and the sharing of science. *Proceedings of the*
457 *National Academy of Sciences of the United States of America*. 2014;**111** Suppl 4(Suppl 4):13642-9.
458 10.1073/pnas.1317511111.
- 459 22. Nagaraj A, Shears E, de Vaan M. Improving data access democratizes and diversifies science.
460 *Proceedings of the National Academy of Sciences of the United States of America*.
461 2020;**117**(38):23490-8. 10.1073/pnas.2001682117.
- 462 23. Roche DG. Evaluating Science's open-data policy. *Science*. 2017;**357**(6352):654.
463 10.1126/science.aan8158.
- 464 24. Hamilton DG, Fraser H, Fidler F, McDonald S, Rowhani-Farid A, Hong K, et al. Rates and
465 predictors of data and code sharing in the medical and health sciences: Protocol for a systematic review
466 and individual participant data meta-analysis. *F1000Research*. 2021;**10**:491.
467 10.12688/f1000research.53874.2.
- 468 25. Tannenbaum S, Ross JS, Krumholz HM, Desai NR, Ritchie JD, Lehman R, et al. Early
469 Experiences With Journal Data Sharing Policies: A Survey of Published Clinical Trial Investigators.
470 *Annals of internal medicine*. 2018;**169**(8):586-8. 10.7326/m18-0723.
- 471 26. Waithira N, Mutinda B, Cheah PY. Data management and sharing policy: the first step towards
472 promoting data sharing. *BMC medicine*. 2019;**17**(1):80. 10.1186/s12916-019-1315-8.
- 473 27. Serwadda D, Ndebele P, Grabowski MK, Bajunirwe F, Wanyenze RK. Open data sharing and the
474 Global South—Who benefits? *Science*. 2018;**359**(6376):642-3. doi:10.1126/science.aap8395.
- 475 28. Forero DA, Curioso WH, Patrinos GP. The importance of adherence to international standards for
476 depositing open data in public repositories. *BMC Research Notes*. 2021;**14**(1):405. 10.1186/s13104-
477 021-05817-z.
- 478 29. Kidwell MC, Lazarević LB, Baranski E, Hardwicke TE, Piechowski S, Falkenberg L-S, et al.
479 Badges to Acknowledge Open Practices: A Simple, Low-Cost, Effective Method for Increasing
480 Transparency. *PLOS Biology*. 2016;**14**(5):e1002456. 10.1371/journal.pbio.1002456.
- 481 30. Riedel N, Kip M, Bobrov E. ODDPub – a Text-Mining Algorithm to Detect Data Sharing in
482 Biomedical Publications. *Data Sci J*. 2020;**19**(1):42. <https://doi.org/10.5334/dsj-2020-042>.
- 483