

ORIGINAL ARTICLE

Many researchers were not compliant with their published data sharing statement: a mixed-methods study

Mirko Gabelica^a, Ružica Bojčić^b, Livia Puljak^{c,*}

^aDepartment for otorhinolaryngology, with head and neck surgery, University Hospital Centre Split, Spinčićeva 1, 21000, Split, Croatia

^bInstitute of Emergency Medicine of Karlovac County, Ul. Dr. Vladka Mačeka 48, 47000, Karlovac, Croatia

^cCenter for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Ilica 242, 10000, Zagreb, Croatia

Accepted 24 May 2022; Published online 30 May 2022

Abstract

Objectives: The objective of the study was to analyze researchers' compliance with their data availability statement (DAS) from manuscripts published in open-access journals with the mandatory DAS.

Study Design and Setting: We analyzed all articles from 333 open-access journals published during January 2019 by BioMed Central. We categorized types of the DAS. We surveyed corresponding authors who wrote in the DAS that they would share the data. Consent to participate in the study was sought for all included manuscripts. After accessing raw data sets, we checked whether data were available in a way that enabled reanalysis.

Results: Of 3556 analyzed articles, 3416 contained the DAS. The most frequent DAS category (42%) indicated that the data sets are available on reasonable request. Among 1792 manuscripts in which the DAS indicated that authors are willing to share their data, 1669 (93%) authors either did not respond or declined to share their data with us. Among 254 (14%) of 1792 authors who responded to our query for data sharing, only 123 (6.8%) provided the requested data.

Conclusion: Even when authors indicate in their manuscript that they will share data upon request, the compliance rate is the same as for authors who do not provide the DAS, suggesting that the DAS may not be sufficient to ensure data sharing. © 2022 Elsevier Inc. All rights reserved.

Keywords: Data availability statement; Data sharing; Open data; Noncompliance; Metaresearch; Reproducibility

1. Introduction

Research data are any information that has been collected within a study to validate original research findings [1]. Research data can be shared with third parties on request or can be made openly available to the public. According to the Open Data Institute, “Data sharing is

providing restricted data to restricted organisations or individuals,” while “Open data is providing unrestricted data to everyone” [2].

A spirit of openness of research data is being advocated in the research community. It is considered that such openness can foster research reproducibility and replicability and enable new discoveries using existing data [3]. However, multiple studies have shown that researchers are not willing to share their data on request [4–6].

In 2019, we reported that few randomized controlled trials' (RCTs) authors were willing to share their data on request. Furthermore, few trials contained a data availability statement (DAS) suggesting a willingness to share trial data. Even when such a DAS was present in the article, very few authors complied with their own DAS and shared the data [7]. However, that study [7] analyzed articles from journals that did not have a mandatory requirement that authors need to report the DAS in the manuscript.

Declarations: Availability of data and materials. We did not publish our raw data along with the manuscript because it could be understood that we are publicly shaming authors who did not want to share their data. As for the raw data that were received during this study, we informed our study participants that those raw data will be deleted after being examined and that all data and communication will be treated with strict confidence.

Competing interests: The authors do not have any conflict of interest to declare.

Funding: This study did not receive extramural funding.

* Corresponding author. Tel.: +385-0-1-370-66-00; fax: +385-0-1-370-66-01.

E-mail addresses: livia.puljak@gmail.com; livia.puljak@unicath.hr (L. Puljak).

What is new?**Key findings**

- We analyzed 3416 articles published by BioMed Central that contained a data availability statement (DAS); the most frequent DAS category (42%) indicated that the data sets are available on reasonable request.
- Of 1792 manuscripts in which the DAS indicated that authors are willing to share their data, only 123 (6.8%) provided the requested data.

What this adds to what is known?

- It was known already that authors might not be willing to share their data; this study shows that they are not willing to do so, even if they write in their DAS that they will share their data on request.

What is the implication and what should change now?

- The authors do not behave in line with their DAS; thus, having a DAS is not a guarantee that the authors will share their data.
- Interventions and incentives for fostering open data and data sharing are needed.

Thus, the present study aimed to analyze researchers' compliance with their DAS from manuscripts published in open-access journals with the mandatory DAS.

2. Methods*2.1. Study design*

We conducted a mixed-methods study, including a methodological study of literature and a cross-sectional survey of corresponding authors.

2.2. Ethics

The Ethics Committee of the University of Split School of Medicine approved the study protocol. Scanned approval of the Ethics Committee was available to the study participants on request. Study participants were informed in the invitation that their response to the e-mailed questions is considered as their written informed consent.

2.3. Journal and manuscript eligibility criteria

We checked for eligibility all manuscripts published during January 2019 in all open-access journals from BioMed

Central (BMC; part of Nature Springer). On the BMC webpage, there were 333 such journals in March 2020. We included in the analysis full reports (i.e., not conference abstracts) of research reports that had a DAS. Any type of research report was eligible, that is, there were no restrictions in terms of study design.

Two authors (M.G. and R.B.) independently conducted the study eligibility check by the manual screening of each article. The third author (L.P.) additionally verified the studies suggested for exclusion. We noted reasons for the exclusion of records.

2.4. Data extraction

We first extracted the following data: journal name, article title, and availability of data and materials. Data were extracted using a computer web scraping tool designed for this study (available at: <https://github.com/bojicicm/bmc-scraper>). In parallel, we extracted data from one thousand articles manually to verify the computer-extracted findings. For the manual extraction, we started with the journals in an alphabetical order. In each journal, we accessed articles in the order from the oldest published to the more recent ones. After opening each manuscript, we chose "Availability of data and materials" in the menu titled "Sections." We manually copied the verbatim text from this section of the manuscript and pasted the text into the electronic spreadsheet. Computer-extracted data were in complete concordance with the data extracted manually.

From manuscripts in eligible journals, we extracted with the web scraper the following data: journal name, International Standard Serial Number, Digital Object Identifier, article title, DAS copied verbatim, and DAS category. We used six categories of the DAS according to the Springer Nature DAS guidance for authors and editors [8]. Those categories were as follows:

- The authors have indicated in which repository they deposited data sets, and they should provide a web link to the data sets.
- The data sets are available from the corresponding author on reasonable request.
- All data generated or analyzed are included in this published article and its supplementary information files.
- The data sets generated are not publicly available due to disclosed reasons but are available from the corresponding author on reasonable request.
- Data sharing does not apply to this article because no data sets were generated or analyzed during this study.
- The data are available from a third party, and restrictions apply regarding data availability because data were used under license and therefore are not publicly available. Data are, however, available upon reasonable request and with the permission of the license holder [8].

For each manuscript where the scraping tool left an empty field for the DAS (indicating that a DAS was not present), we checked manually whether indeed a DAS was not reported in the manuscript.

For manuscripts that were not eligible for classification according to Springer Nature DAS guidance for authors and editors, we created a seventh and eighth category: 7. not available—for statements that claim data are not available to the third party under any circumstances and 8. other—for statements that cannot be classified as categories one to seven. Some manuscripts had dual or triple coding because the DAS had elements of several categories.

Articles with the DAS that fell into multiple categories were eligible for inclusion in author survey if they included DAS categories 2, 4, or 6. For the author survey, we extracted data on the corresponding author name and corresponding author email. We extracted data into a Microsoft Excel worksheet (Microsoft Inc., Redmond, WA, USA). Two authors piloted the worksheet on a sample of twenty articles to ensure that it is suitable to extract target data.

2.5. Author survey

All corresponding authors of manuscripts with DAS categories 2, 4, or 6 were contacted via e-mail and asked to share their raw data sets. A deidentified copy of the e-mail sent to the authors is available as [Supplementary File 1](#). We have also prepared a nondisclosure agreement (NDA) for researchers who might request it ([Supplementary File 2](#)). The first author (M.G.) sent personalized e-mails to each corresponding author from an official hospital employee e-mail account. If the authors did not respond after the initial e-mail, they received only one reminder. If the authors responded positively with a willingness to share raw data sets but did not provide data within 2 weeks, they received an additional e-mail reminder. Corresponding e-mails were obtained directly from manuscripts included in the first part of the study. All e-mails and initial reminders were sent between January 18th and May 18th in 2021.

If corresponding authors suggested we should contact another team member to obtain data and provided their e-mail addresses, we contacted those persons. If the authors indicated that additional regulatory or approval procedures were required for obtaining raw data sets, we did engage in those processes, such as signing an NDA or data transfer agreements (DTAs) or sending an official letter of request signed by the University of Split School of Medicine official. If the message sent to the corresponding authors was returned undelivered, we did not find their alternative e-mail address. If the corresponding authors did not respond, we did not attempt to contact other co-authors.

After accessing raw data sets, we checked whether data were available in a way that enabled reanalysis, that is, published in a file that enabled data use and whether relevant metadata were included.

2.6. Data analysis

We presented descriptive data as frequencies and percentages. We also performed a qualitative content analysis of the responses to the survey, from the authors who were unwilling to share their data. We used the inductive approach, examining patterns and concepts in a set of data, to report reasons for not sharing the data.

3. Results

3.1. Data extraction analysis

In January 2019, BMC had 333 journals listed on the web page. Among those journals, 51 (15%) did not publish any articles in the analyzed period, while 282 (84%) published at least one article. We extracted 3556 articles from those 282 journals. We excluded 68 articles, including 63 corrections, and five conference reports. Among the remaining 3488 records, we further excluded 72 articles that did not have a DAS. Web scraper software has found 72 articles without a DAS; all of them were manually checked, and no discrepancy was found between software and human verification. Thus, we included in our analysis 3416 articles reporting original research that had a DAS ([Fig. 1](#)).

The most used DAS category was category #2 (the data sets are available from the corresponding author on reasonable request), followed by category #3 (all data generated or analyzed are included in this published article and its supplementary information files). The minority of the articles were classified into multiple categories ([Table 1](#)).

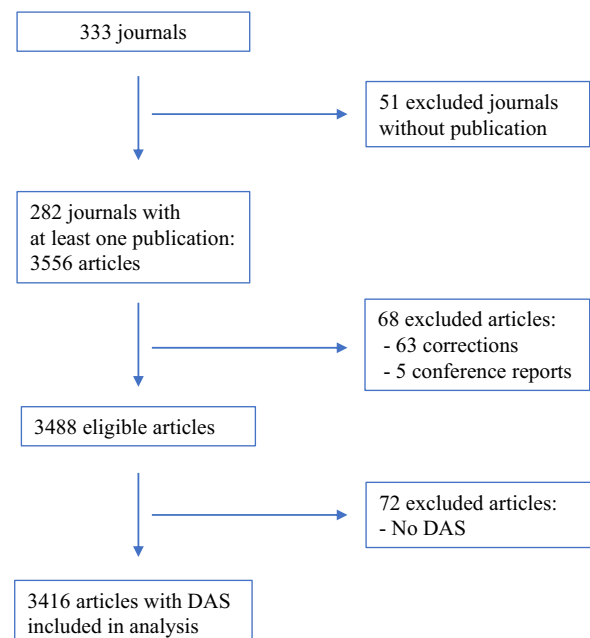


Fig. 1. Article inclusion flowchart. DAS, data availability statement.

Table 1. Frequency of different categories of the data availability statement (DAS) ($N = 3416$)

DAS category	N (%)
1. The authors have indicated in which repository they deposited data sets, and they should provide a web link to the data sets	369 (10.80)
2. The data sets are available from the corresponding author on reasonable request.	1415 (41.42)
3. All data generated or analyzed are included in this published article and its supplementary information files.	571 (16.71)
4. The data sets generated are not publicly available due to disclosed reasons but are available from the corresponding author on reasonable request.	159 (4.65)
5. Data sharing does not apply to this article because no data sets were generated or analyzed during this study.	416 (12.17)
6. The data are available from the third party; restrictions apply regarding data availability because data were used under license and therefore are not publicly available. Data are, however, available upon reasonable request and with permission of the license holder	121 (3.54)
7. Not available—for statements that claim data are not available to the third party under any circumstances	89 (2.60)
8. Other—for statements that cannot be classified as categories one to seven	122 (3.57)
DAS categorized into two categories	152 (4.45)
DAS categorized into three categories	2 (0.05)

Among 129 articles with the DAS categorized into more than one category, the most common combination category of the DAS was #1 and #3 (Table 2).

3.2. Author survey

There were 1792 manuscripts (category #2: 1415, category #4: 159, category #6: 121, dual-coded containing categories #2, #4 or #6: 95, and triple-coded containing category #2: 2) eligible for the author survey. We contacted all the 1792 corresponding authors from the eligible manuscripts to request their data. After our initial e-mail, we received no

Table 2. Frequency of data availability statements categorized into more than one category ($N = 154$)

DAS categories	N (%)
1 and 2	11 (7.14)
1 and 3	51 (33.11)
1 and 8	(0.64)
2 and 3	55 (35.71)
2 and 4	5 (3.24)
2 and 5	2 (1.29)
2 and 6	2 (1.29)
2 and 7	4 (2.59)
2 and 8	4 (2.59)
3 and 4	4 (2.59)
3 and 7	1 (0.64)
3 and 8	3 (1.94)
4 and 6	4 (2.59)
4 and 7	1 (0.64)
6 and 7	2 (1.29)
6 and 8	1 (0.64)
7 and 8	1 (0.64)
1 and 2 and 3	2 (1.29)

DAS, data availability statement.

reply from 1416 (79%) contacts, and 77 (4.3%) e-mails bounced back as undelivered. There were 38 corresponding authors who instructed us to contact another researcher responsible for data management, retrieval, and sharing and provided us with a forwarding address; 29 of those contacts did not respond. From 17 researchers, we received an automated e-mail stating they were unavailable at the time, and we should contact them after the “away” period, which we did; only one researcher responded to the reminder e-mail sent after the “away” period.

In summary, of 1792 e-mails sent, we did not receive any response for 1538 articles because messages were not delivered ($N = 77$; 4.3%) or the author did not reply ($N = 1461$; 81%). Responses were received from 254 (14%) contacted authors. A flow chart presenting the outcome of author contacts is shown in Figure 2.

3.3. Data sharing

Among the 254 authors who responded to our e-mail, 123 shared the data, which corresponds to 7.1% of 1715 articles for which authors were contacted, and e-mails did not bounce back undelivered. Reasons for not sharing data among the remaining 131 authors who responded to our request for data are provided in Table 3.

Thirty authors asked for more information about the study. We replied to them all, and the clarifications sometimes took several e-mails. Of those 30 authors, seven eventually provided their data.

Among 22 authors who requested that we sign an NDA or a DTA, two authors accepted the NDA that we have prepared, while 20 sent their version of the NDA or DTA. We received data sets from eight of those 22 authors. Ten authors did not reply at all after we sent them an NDA or DTA. Two authors had requests regarding the NDA that we could not accommodate; namely, one author wanted an NDA signed by the principal (rector) of the University of Split, and another one wanted the NDA signed by an official from the university

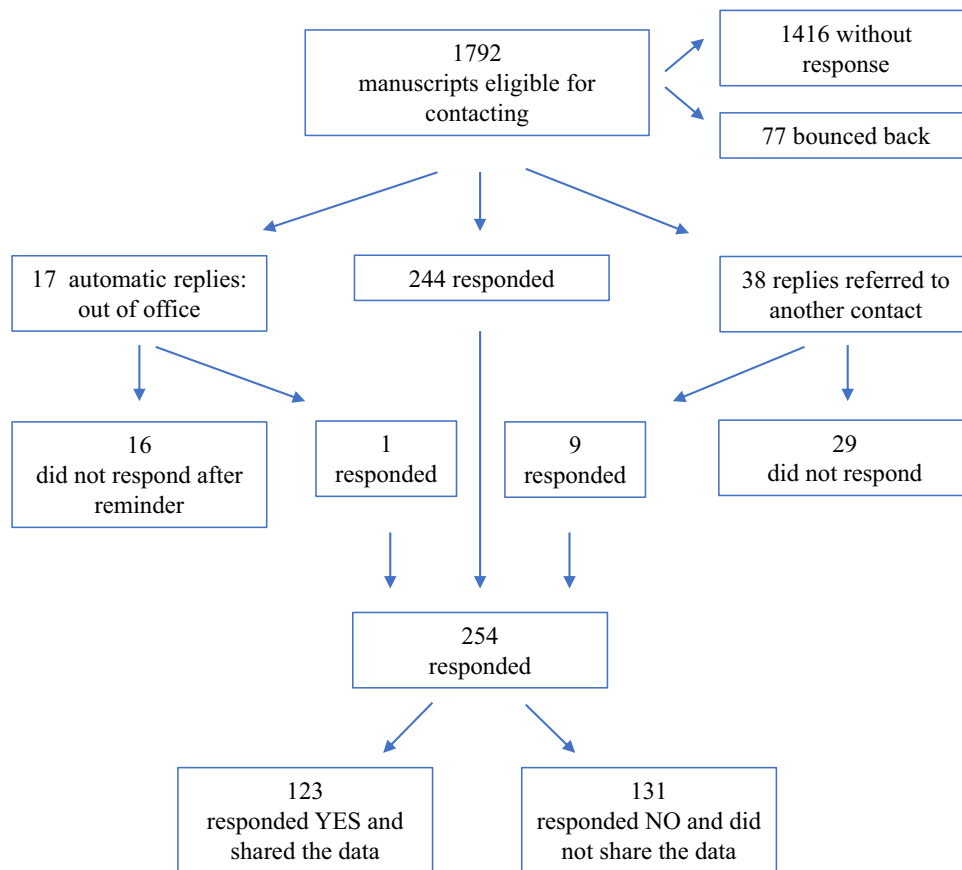


Fig. 2. Flow chart presenting the outcome of author contacts in the survey of authors.

technology transfer office. One author informed us that data sharing is against their ethical committee suggestions after both parties signed the NDA (Table 4).

Two authors demanded reimbursement, and one author requested co-authorship for providing us with data from their research. Eight researchers directed us to a web portal we should access and register, with instructions that afterward we should specifically describe variables we need and only then would their decision-making start. We did not engage in those activities. Two authors asked us to send them the official letter, on the school letterhead, in which we will request the data. One of those authors shared the data after sending the official letter, while the other did not. Various other reasons for not sharing the data, including health condition and retirement of the author, requests for translating the ethics approval for our research into non-English language, large data sets, misplacing the entire data from the study, and so on, are shown in Table 3.

3.4. Study design of articles whose authors shared the data

Based on the authors' self-report, the most common study design of the articles for which authors shared raw data was cross-sectional (Table 4). Among the 122 articles

for which the authors shared the raw data, there were 11 (9%) RCTs.

3.5. Useable data sets

Among 123 data sets shared with us, 117 (95%) were useable, that is, they were sent in a format that would allow reanalysis of the data. The five unusable data sets were sent in a Portable Document Format file.

4. Discussion

Among authors whose DAS indicated data are available on request, we received responses from 14% of the contacted authors, and data were shared by 6.8% of the contacted authors. The percentage of useable data sets was 6.7%.

Among the contacted authors in both studies, some provided reasons for not sharing data that offer insight into hurdles that individuals requesting data may face when accessing raw data from published articles. Ventresca et al. [5] addressed incentives that could be offered to authors in exchange for data. They suggest that individuals who made considerable effort should be recognized by offering

Table 3. Reasons for not sharing requested data (*N* = 132)

Reason	<i>N</i> (%)
The authors asked for more information about our study, but after our detailed response and clarification, we did not receive further response from them	23 (17.42)
Their informed patient consent did not include sharing data with other researchers, or the ethical committee prohibited external data sharing and use	14 (10.6)
They cannot access the data, either because they are no longer in the institution that conducted the research or they are no longer active on the project	12 (9.09)
They do not want to share the data or in any way participate in our study without a specific explanation	11 (8.33)
No reply after we signed and sent an NDA or DTA that the authors requested	10 (7.57)
The corresponding author instructed us to use a web service to request the data	8 (6.06)
The data do not belong to them	6 (4.54)
The study was still ongoing	5 (3.78)
Privacy concerns; specifically, they did not want to share deidentified patient data, and other concern was third party data storage and safekeeping from external access to patient data	4 (3.03)
This was a summary article, and there are no data to share	3 (2.27)
We were unable to meet their specific requests regarding the NDA	2 (1.51)
The author requested reimbursement for data sharing	2 (1.51)
The author wanted ethics approval translated into English	2 (1.51)
I will send you the data (however, the authors did not send the data subsequently)	2 (1.51)
After both parties signed the NDA, the author wrote back to say that data sharing is against their ethical committee suggestions	1 (0.75)
A conceptual article, no data to share	1 (0.75)
The data set has 8 Gb and is unable to share such a large file	1 (0.75)
The most important data and the data set is in the supplementary file in the manuscript (however, this was not the case; the supplementary file did not contain raw data from the study)	1 (0.75)
The author misplaced the data	1 (0.75)
The author was not sure what data we needed (the response remained the same even after we sent three emails to explain what information we needed)	1 (0.75)
Repeat your inquiry in 6 months	1 (0.75)
The author was sick and did not go to the office	1 (0.75)
Another scientist will send you the data (but they did not)	1 (0.75)
The author wanted to schedule an online meeting but was unable to set the date and time regarding the time zone difference	1 (0.75)
The author wrote they would get back to us (but did not)	1 (0.75)
The author requested vast and lengthy procedures and authorizations	1 (0.75)
I will respond to your message in due course	1 (0.75)
I am writing a grant application and cannot help you	1 (0.75)
The author sent us published supplementary materials	1 (0.75)
“We did not use any of our own data for this publication. The data were gleaned from peer-reviewed publications that are widely available and referenced in the reference section.”	1 (0.75)
Qualitative research conducted in Ukrainian and Russian language	1 (0.75)
“What advantage will I get from sharing?”	1 (0.75)
“Too many technical aspects from our side; cannot comply with.”	1 (0.75)
Qualitative research conducted in the Finnish language	1 (0.75)
Awaiting ethical committee approval	1 (0.75)
It will take too long	1 (0.75)
All data are within the article (however, the data were not within the article)	1 (0.75)
The author asked if our institution has MATLAB software; no reply after two reminders	1 (0.75)
The author wanted us to sign their DTA, which we agreed to. However, they did not send the DTA or replied to our messages even after we sent them two reminders.	1 (0.75)

(Continued)

Table 3. Continued

Reason	N (%)
Wants that NDA is signed by the rector of the University of Split	1 (0.75)
Do not want to prepare all this for a study that is not interested in the data itself	1 (0.75)
It may be not feasible to do this in the near future	1 (0.75)

NDA, nondisclosure agreement; DTA, data transfer agreement.

co-authorship to the original study. However, they needed those data for their individual patient data (IPD) meta-analysis [5].

In our study, we did not plan a priori to offer co-authorship as we did not aim to reuse the data and publish such new analysis. Offering co-authorship for sharing data has been questioned as potentially unethical. For example, Devriendt et al. warn that although co-authorship in return for providing data is expected, this might not be compatible with the internationally recognized authorship guidelines and that, furthermore, raises concerns over the ability of secondary analysts to potentially contest the proposed research methods or conclusions that were initially drawn from the data [9].

In our study, we did not offer any reimbursement to the authors for their efforts related to data sharing. Ventresca et al. described that they tried an approach of offering reimbursement for minimal expenses related to data sharing, for example, shipping fees for data that the corresponding authors did not want to send electronically [5]. Veroniki et al. tried

Table 4. Study design of articles for which the authors shared the raw data, as reported in the manuscript ($N = 122$)

Study design	N (%)
Not reported	47 (38.52)
Cross-sectional	25 (20.49)
Randomized controlled trial	11 (9.01)
Cohort	6 (4.91)
Retrospective	5 (4.09)
Mixed methods	4 (3.27)
Systematic review	4 (3.27)
Prospective study	3 (2.45)
Protocol	3 (2.45)
Qualitative study	3 (2.45)
Questionnaire study	2 (1.63)
Descriptive study	2 (1.63)
Observational study	1 (0.81)
Post hoc analysis	1 (0.81)
Case-control study	1 (0.81)
Pre-test/post-test	1 (0.81)
Meta-analysis	1 (0.81)
Interventional study	1 (0.81)
Case study	1 (0.81)
Total	122

offering a small financial incentive of 100 Canadian dollars to the authors of trials eligible for an IPD meta-analysis, but this intervention did not improve IPD retrieval rates [10,11].

Some authors wanted us to log into specific web platforms, with complicated procedures involved as a prerequisite for them to start considering our data request. However, we did not engage in those processes, as our prior experience and several other manuscripts indicate that such a decision process is often lengthy and ultimately with a negative outcome [7,12].

Several authors did not share their data with us with the explanation that they conducted qualitative studies. Due to fundamental differences in qualitative and quantitative studies, it has been reported that qualitative studies warrant specific considerations in the data-sharing movement [13]. Unlike the set of numbers expected to be shared for quantitative studies, to enable statistical reanalysis, data in qualitative studies are usually collected via interviews, focus groups, direct observation, and document review. These differences between quantitative and qualitative studies may have repercussions on the reproducibility of results. Reproducibility is defined as obtaining a consistent result by “using the same input data; computational steps, methods, and code; and conditions of analysis,” thus implying computational reproducibility [14].

This idea of computational verification, that is, reproducibility, may not translate well to qualitative studies and whether interview transcripts can genuinely be considered raw data [13,14]. There are also language hurdles with sharing data from interviews [14]. Indeed, in our study, some authors explained that they do not wish to share their data from qualitative studies because interviews were conducted in Ukrainian and Russian languages. However, the language barrier should not be a priori reason for not sharing such data because individuals requesting data could have the necessary language proficiency or resources to secure translation.

The availability of raw data, even on request, is considered as a safeguard of good research practices. An editor of the journal *Molecular Brain* published an editorial in 2020 describing his effort to request raw data from manuscripts. Since 2017, he has requested raw data from 41 manuscripts. To his surprise, the authors of 21 (51%) of those 41 manuscripts decided to withdraw their manuscript without providing raw data. The editor rejected 19 out of 20 remaining manuscripts because of insufficient raw data. Thus, the editor hypothesized that either raw data did not exist at all or at least in some portions. The editor

concluded that journals should request raw data from the authors in order to verify possible data fabrication, increase research results' reproducibility, and strengthen public trust in science [15].

It is possible that some of the authors that we have contacted did not respond to our request because they do not have raw data or because of problems with their raw data.

There is much emphasis on data sharing and open data currently in the research community. However, even though many authors express support for ideas of data sharing and open data [4,16,17], when it comes to sharing data, the authors may not behave in line with what they say [7,18].

Authors' reluctance to share research data has direct implications for clinical medicine on several levels. The availability of data enables reanalyses and ensures reproducibility of the published data. Ebrahim et al. explored reanalyses of clinical trial data and showed that 35% of those reanalyses could not confirm data reported in the original publication [19]. We also checked whether the shared data would enable reanalysis, but we did not attempt to reanalyze the data, as this was not the aim of this study. Furthermore, data availability may safeguard against research misconduct. For example, in early 2020, following the emergence of COVID-19, two studies published in prominent medical journals were retracted due to concerns about fraudulent data [20]. In August 2021, Lee et al. reported that one of those studies was still widely cited, despite its early retraction [20]. When it comes to an emergent public health crisis, such as COVID-19, it could be expected that authors would be more prone to data sharing and open data. However, Strcic et al. reported that raw data and/or code were found in 15% of all preprint articles that were posted on bioRxiv and medRxiv in the first months of the COVID-19 pandemic [21].

Also, IPD meta-analyses are not possible without the availability of IPD, and the authors of such meta-analyses need to rely on the authors of primary studies for data sharing. Scut et al. reported their experience in requesting IPD from trials in vascular and cognitive medicine for their meta-analysis. They received IPD from 25% of the authors whose e-mail address was found [22].

Currently, there are no mandatory requirements related to data sharing. However, data sharing statements are being widely adopted. The International Committee of Medical Journal Editors (ICMJE) postulates that reports of clinical trials published from July 2018 onward need to provide a mandatory data sharing plan—but this applies only to clinical trials, not to other types of studies [23]. Nevertheless, there is no guarantee that a data sharing plan will lead to actual data sharing. A proof of this is the study by Danchev et al., who analyzed 487 trials published in JAMA, Lancet, and New England Journal of Medicine after implementation of the ICMJE data sharing statement requirement. There were 69% of articles where the authors declared data sharing. Only 2 (0.6%) IPD sets were actually deidentified and publicly available on a journal website. Of the 89

articles where the authors reported that IPD would be stored in secure repositories, data from only 17 articles were found in the respective repositories by April 10, 2020 [24].

One solution for solving some of the problems with data availability was proposed by Wolfenden et al., who suggested that data extracted from trials included in systematic reviews should be shared with the research community [25].

Based on our findings, requesting a DAS is not enough. Editors as gatekeepers should start requesting raw data as the obligatory part of manuscript submission; this could likely be the only way to secure the accessibility and verifiability of raw data from published studies.

A limitation to this study could be its inherent aim. We did not intend to conduct reanalysis of the data or to do follow-up studies on the raw data. It is possible that the corresponding authors could respond differently if our study aim was related to their data. Another limitation is that we sought consent to participate in the study by simply replying to the email we have sent. We received responses from 254 individuals; this could be considered as 254 individuals who consented to participate in the study. It is possible that some authors did not want to reply or ask for clarifications regarding the study because they thought that would automatically mean they consented to data sharing. Furthermore, in our invitation e-mail, we indicated that our team is interested in re-examining those raw data sets, whether or not they are adequate for reanalysis. Without further clarifications, for example, information about the research context, that a lot of research findings have been found to not reproduce on reanalysis, it is possible that our invitation was not sufficiently compelling to the targeted individuals. Thus, it is possible that different wording of the invitation could have yielded more responses from the invited individuals. Nevertheless, we consider that all the contacted individuals were free to contact us to request more information about the study and our intentions, but few have done so.

In this study, we used articles published in the BMC journal. For our sample, we were looking for a publisher with a large volume of journals/articles and with a mandatory DAS. However, it is possible that results obtained using the BMC journals may not be generalizable for all research articles.

In conclusion, authors of research articles are frequently not willing to share their data, even if they wrote in their manuscript DAS that they would do so. Our findings can enable the creation of new guidelines and practices in the research community to foster the availability of raw research data.

CRedit authorship contribution statement

Mirko Gabelica: Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review & editing. **Ružica Bojčić:** Data curation, Formal analysis, Investigation, Writing – original draft, Writing – review

& editing. **Livia Puljak:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Project administration, Supervision, Writing – original draft, Writing – review & editing.

This study was part of the PhD thesis of the first author (M.G.). The thesis was defended at the University of Split School of Medicine, Croatia. The authors are grateful to all study participants for taking part in the study.

Supplementary data

Supplementary data to this article can be found online at <https://doi.org/10.1016/j.jclinepi.2022.05.019>.

References

- [1] University of Leeds. What is research data? Research data management explained. Available at https://library.leeds.ac.uk/info/14062/research_data_management/61/research_data_management_explained. Accessed April , 2022.
- [2] Tenninson J. Data sharing is not open data. Available at <https://www.governmentcomputing.com/central-government/features/featured-data-sharing-is-not-open-data-4198712>. Accessed April , 2022.
- [3] Gewin V. Data sharing: an open mind on open data. *Nature* 2016;529:117–9.
- [4] Borghi JA, Van Gulick AE. Data management and sharing in neuroimaging: practices and perceptions of MRI researchers. *PLoS One* 2018;13:e0200562.
- [5] Ventresca M, Schunemann HJ, Macbeth F, Clarke M, Thabane L, Griffiths G, et al. Obtaining and managing data sets for individual participant data meta-analysis: scoping review and practical guide. *BMC Med Res Methodol* 2020;20:113.
- [6] Wu IXY, Xiao F, Wang H, Chen Y, Zhang Z, Lin Y, et al. Trials number, funding support, and intervention type associated with IPDMA data retrieval: a cross-sectional study. *J Clin Epidemiol* 2021;130:59–68.
- [7] Gabelica M, Cavar J, Puljak L. Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *J Clin Epidemiol* 2019;109:111–6.
- [8] Springer Nature. Data availability statements. Available at <https://www.springernature.com/gp/authors/research-data-policy/data-availability-statements/12330880>. Accessed April 5, 2022.
- [9] Devriendt T, Shabani M, Borry P. Data sharing in biomedical sciences: a systematic review of incentives. *Biopreserv Biobank* 2021;19:219–27.
- [10] Veroniki AA, Ashoor HM, Le SPC, Rios P, Stewart LA, Clarke M, et al. Retrieval of individual patient data depended on study characteristics: a randomized controlled trial. *J Clin Epidemiol* 2019;113:176–88.
- [11] Veroniki AA, Straus SE, Ashoor H, Stewart LA, Clarke M, Tricco AC. Contacting authors to retrieve individual patient data: study protocol for a randomized controlled trial. *Trials* 2016;17:138.
- [12] Puljak L, Marin A, Vrdoljak D, Markotic F, Utrobovic A, Tugwell P. Celecoxib for osteoarthritis. *Cochrane Database Syst Rev* 2017;5:CD009865.
- [13] Tsai AC, Kohrt BA, Matthews LT, Betancourt TS, Lee JK, Papachristos AV, et al. Promises and pitfalls of data sharing in qualitative research. *Soc Sci Med* 2016;169:191–8.
- [14] National Academies of Sciences Engineering and Medicine. Reproducibility and replicability in science. Washington, DC: National Academies Press; 2019:39–54.
- [15] Miyakawa T. No raw data, no science: another possible source of the reproducibility crisis. *Mol Brain* 2020;13:24.
- [16] Hrynaskiewicz I, Harney J, Cadwallader L. A survey of researchers' needs and priorities for data sharing. *Data Sci J* 2021;20(1):31.
- [17] Houtkoop BL, Chambers C, Macleod M, Bishop DVM, Nichols TE, Wagenmakers EJ. Data sharing in psychology: a survey on barriers and preconditions. *AMPPS* 2018;1:70–85.
- [18] Bergeris A, Tse T, Zarin DA. Trialists' intent to share individual participant data as disclosed at ClinicalTrials.gov. *JAMA* 2018;319:406–8.
- [19] Ebrahim S, Sohani ZN, Montoya L, Agarwal A, Thorlund K, Mills EJ, et al. Reanalyses of randomized clinical trial data. *JAMA* 2014;312:1024–32.
- [20] Lee TC, Senecal J, Hsu JM, McDonald EG. Ongoing citations of a retracted study involving cardiovascular disease, drug therapy, and mortality in COVID-19. *JAMA Intern Med* 2021;181:1535–7.
- [21] Strcic J, Civljak A, Gloznic T, Pacheco RL, Brkovic T, Puljak L. Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv. *Scientometrics* 2022;127(5):2791–802.
- [22] Scutt P, Woodhouse LJ, Montgomery AA, Bath PM. Data sharing: experience of accessing individual patient data from completed randomised controlled trials in vascular and cognitive medicine. *BMJ Open* 2020;10:e038765.
- [23] International Committee of Medical Journal Editors (ICMJE). Clinical trials. Data sharing. Available at <https://www.icmje.org/recommendations/browse/publishing-and-editorial-issues/clinical-trial-registration.html#two>. Accessed April 5, 2022.
- [24] Danchev V, Min Y, Borghi J, Baiocchi M, Ioannidis JPA. Evaluation of data sharing after Implementation of the international committee of medical journal editors data sharing statement requirement. *JAMA Netw Open* 2021;4:e2033972.
- [25] Wolfenden L, Grimshaw J, Williams CM, Yoong SL. Time to consider sharing data extracted from trials included in systematic reviews. *Syst Rev* 2016;5:185.