




# Open data and data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv

Josip Strcic<sup>1</sup> · Antonia Civljak<sup>2</sup> · Terezija Glozinic<sup>1</sup> · Rafael Leite Pacheco<sup>3</sup> · Tonci Brkovic<sup>4</sup> · Livia Puljak<sup>1</sup> 

Received: 27 September 2021 / Accepted: 9 March 2022 / Published online: 25 March 2022  
© Akadémiai Kiadó, Budapest, Hungary 2022

## Abstract

This study aimed to analyze the content of data availability statements (DAS) and the actual sharing of raw data in preprint articles about COVID-19. The study combined a bibliometric analysis and a cross-sectional survey. We analyzed preprint articles on COVID-19 published on medRxiv and bioRxiv from January 1, 2020 to March 30, 2020. We extracted data sharing statements, tried to locate raw data when authors indicated they were available, and surveyed authors. The authors were surveyed in 2020–2021. We surveyed authors whose articles did not include DAS, who indicated that data are available on request, or their manuscript reported that raw data are available in the manuscript, but raw data were not found. Raw data collected in this study are published on Open Science Framework (<https://osf.io/6ztec/>). We analyzed 897 preprint articles. There were 699 (78%) articles with Data/Code field present on the website of a preprint server. In 234 (26%) preprints, data/code sharing statement was reported within the manuscript. For 283 preprints that reported that data were accessible, we found raw data/code for 133 (47%) of those 283 preprints (15% of all analyzed preprint articles). Most commonly, authors indicated that data were available on GitHub or another clearly specified web location, on (reasonable) request, in the manuscript or its supplementary files. In conclusion, preprint servers should require authors to provide data sharing statements that will be included both on the website and in the manuscript. Education of researchers about the meaning of data sharing is needed.

**Keywords** SARS-CoV-2 · COVID-19 · Coronavirus · Preprint server · Open data · Data sharing

## Background

To enable transparency, reproducibility, and conduct of new studies with existing data, it would be important that authors of biomedical studies support the principles of ‘open data’ and data sharing’, and make their raw data publicly available (Watson, 2015). Some

---

✉ Livia Puljak  
livia.puljak@gmail.com; livia.puljak@unicath.hr

Extended author information available on the last page of the article

journals require authors to make their raw data sets publicly available or to make them available on request (Godlee & Groves, 2012; Groves, 2010), but not all.

Even when the authors write in their data sharing statements that the data collected within the study will be available on request, we have shown that authors of clinical trials frequently ignore such requests for data sharing (Gabelica et al., 2019). Federer et al. have analyzed data availability statements in the journal PLoS One to test whether the authors complied with the PLoS policies that require researchers to share the data underlying their results and publications. The results showed that only about 20% of data availability statements indicated that data were deposited in a repository, which is the preferred method. The authors concluded that more stringent policies may be needed to increase data sharing (Federer et al., 2018). It would be optimal if all the authors would make their raw data publicly available at the time of publication.

The importance of data sharing is particularly compelling in a situation such as the current pandemic of disease COVID-19 caused by the novel coronavirus (Moorthy et al., 2020). Moorthy et al. highlighted in their editorial published in the Bulletin of the World Health Organization [quote]: “*Rapid data sharing is the basis for public health action*” (Moorthy et al., 2020). Open sharing of data is essential during public health emergencies (Homolak et al., 2020; Modjarrad et al., 2016). While it has been acknowledged that conducting research during the pandemic is associated with inherent challenges, it was highlighted that data sharing during the pandemic was of utmost importance (Wolkewitz & Puljak, 2020).

Besancon et al. argued that open science saves lives while criticizing violations of open science principles during the COVID-19 pandemic. They provided evidence of the misuse of open science principles at different stages of the scientific process during the pandemic, which contributed to research waste (Besancon et al., 2021).

Lucas-Dominguez et al. analyzed COVID-19 publications from the early pandemic to evaluate the research data available in publications or deposited in repositories. They searched PubMed Central and found underlying research data available for only 13.6% of the analyzed records. They concluded that data sharing was not a common practice, even in health emergencies, such as the beginning of the COVID-19 pandemic (Lucas-Dominguez et al., 2021). However, such analysis of data availability is not available for preprint articles.

Preprint servers allow early sharing of scientific articles. Such servers enable authors to simply make their manuscripts publicly available before submission/acceptance in a scholarly journal. In this way, scholarly information becomes rapidly available to the public. Beyond enabling rapid dissemination of research methods and findings, the preprints also solicit feedback from the research community to help improve the final manuscript (Iacobucci, 2019). By sharing a manuscript on a preprint server, it is anticipated that the authors may receive feedback from a larger community compared to a typical peer-review that involves comments from two or three experts in the field (Poremski et al., 2019).

However, there are also some challenges associated with preprint articles. As such articles did not go through editorial and peer-review checks (Hoy, 2020), the preprint was described as “interim research product” (Poremski et al., 2019). Furthermore, some preprint servers do not conduct any screening and quality checks. This can lead to the dissemination of misleading information (Pourhoseingholi et al., 2020). It has been described that only three preprint servers, including Research Square, bioRxiv and medRxiv, check if the content contains unfounded medical claims (Kirkham et al., 2020).

During the COVID-19 pandemic, the preprint servers had a prominent role in the dissemination of research results. In the first three months following the COVID-19 outbreak,

there were 533 articles with original data in the World Health Organization (WHO) COVID-19 collection of research articles from scholarly journals, compared with 1088 preprint articles with research data published on bioRxiv and medRxiv preprint servers (Fidahic et al., 2020). However, it is not known how often authors that post their manuscripts in preprint servers also make their raw data available together with a manuscript, particularly in the case of a public health emergency.

This study aimed to analyze data sharing statements and actual data sharing in articles about COVID-19 published in preprint servers medRxiv and bioRxiv.

## Methods

### Protocol and registration

The protocol for this study was defined before the commencement of the study. After approval of the protocol by all planned co-authors, the protocol was published on Open Science Framework (<https://osf.io/6ztec/>) on April 7, 2020.

### Ethics

The study included a survey of corresponding authors of manuscripts deposited in preprint servers. The study protocol was approved by the Ethics Committee of the Catholic University of Croatia (date: April 1, 2020; Klasa: 641-03/20-01/06; Urbroj: 498-03-02-06-02/1-20-02). We began surveying the authors only after we received the approval of the Ethics Committee. All authors consented to participate in the study.

### Eligibility criteria

We included studies about SARS-CoV-2 and COVID-19 published on preprint servers medRxiv and bioRxiv from December 2019 until March 30, 2020. These two preprint servers were chosen because they check submitted articles for unfounded medical claims (Kirkham et al., 2020). Furthermore, they were prominently used by authors in the COVID-19 pandemic (Teixeira da Silva JA, 2021). The last version of the article was analyzed for articles that had multiple versions posted up until the cut-off date of March 30, 2020.

### Search

We located the preprint articles related to COVID-19 by using the hyperlink described in the preprints' homepage as "COVID-19 SARS-CoV-2 preprints from medRxiv and bioRxiv". The medRxiv team provided the information that the decision to add a preprint to the bioRxiv and medRxiv COVID-19/SARS-CoV-2 collection page is determined by their screening team who read the paper and assess whether the content is related to COVID-19/SARS-CoV-2; this decision is not algorithmic or author-initiated (personal communication).

## Data extraction

All articles were exported in JSON format and imported into Excel (Microsoft Inc., Redmond, WA, USA). We designed custom data extraction sheet in Excel, and datasheet was piloted on ten studies by three co-authors (JS, AC, TG) before beginning the extraction of the remaining data. Co-authors' suggestions were taken into account regarding the need for revision of the data extraction table. One author extracted data (JS, AC, TG, RLP participated in data extraction), while two authors (TB, LP) verified data extraction.

For each article, we copied verbatim statements provided in the section “Data/Code”. We recorded if an article did not have any text in the section “Data/Code” on its website. Additionally, we downloaded each pdf file with the full text of the manuscript, extracted name and email address of the corresponding author, and assessed whether the article has included a data sharing statement within the main file; if yes, we copied and categorized the data statement, and we compared statements from the section Data/Code on the article website with data sharing statements within the manuscript file. We also extracted information if the authors reported that they had used data publicly available for their study. We noted discrepancies, such as Data/Code statements reports that data were publicly shared, but the manuscript file did not mention anything regarding data sharing.

For articles that had accompanying shared raw data, we recorded the location of the shared data. We recorded whether we were successful in accessing data. We defined raw data as primary data collected from the source, which has not been processed. There are various types of data that can be collected within a research study. Thus, raw data can be unprocessed facts, raw numbers, figures, images, words or sounds derived from observations or measurements. For example, raw data can include a spreadsheet that contains numbers for various variables collected from participants, or recordings of conversations with participants in a qualitative research study, or another type of data—depending on the nature of the study.

In January 2022, we checked whether the preprint articles included in the analysis were published in a scholarly journal. These data were collected from the included preprint servers. When the manuscript is published in a scholarly journal, this information is prominently displayed below the article on the analyzed preprint servers. The information we extracted were whether the article was published in a journal or not, in which journal, and we tabulated a digital object identifier (DOI) of the article in a scholarly journal.

## Survey of corresponding authors

When we analyzed data sharing statements (DAS), we divided the preprints into 5 categories depending on the need to conduct an author survey: (i) there was no DAS, (ii) data available on request, (iii) manuscript reported that raw data are available in the manuscript, but raw data not found, (iv) DAS unclear, or (v) manuscript not eligible for author survey for any reason, for example, because the authors have transparently reported where their raw data can be found. We surveyed authors whose preprints belonged to one of the first three categories. A template of the email with the survey and accompanying information for participants are available in Supplementary file 1.

We sent an email to all corresponding authors that did not share their raw data publicly with this question: “We would appreciate very much if you could explain us your reasons for not publicly providing raw data collected within your study”. The survey was sent, and

answers were received between May 7, 2020 and May 13, 2021, as some authors provided responses very late.

For authors that have indicated that they will share their data on request, or reasonable request, we also asked them in which circumstances (pre-conditions) they would be willing to share their data on request.

For authors that mentioned in their Data/Code sections, or data sharing statements that data have been shared publicly, but we were not able to locate such data, we asked for clarifications where exactly the data were shared (to exclude the possibility of our misunderstanding).

Each email sent to corresponding authors included information about study authors, approval of the study by the Ethics committee (and approval number), and study protocol, as well as information that their response to the email will be considered as their informed consent. Corresponding authors were informed that they are free to submit any questions to the principal investigator (LP), should they need more information about our study.

## Raw data

All raw data collected within this study are available in Supplementary file 2 and also published on Open Science Framework (<https://osf.io/6ztec/>).

## Synthesis of results

We analyzed data with descriptive statistics and report frequencies and percentages.

## Results

We included in the analysis 897 preprint articles about SARS-CoV-2 and COVID-19 published from January 19, 2020 to March 30, 2020. There were 678 (76%) preprints published on medRxiv and 219 (24%) on bioRxiv.

There were 699 (78%) preprint articles with Data/Code field present on the website of a preprint server. All articles published on medRxiv ( $N=678$ ; 100%) had a Data/Code field, compared to 21 (9.6%) of articles published on bioRxiv.

The ten most common categories of information provided in the Data/Code field on the website are shown in Table 1. All other categories can be seen in Supplementary file 2 (column O). Most commonly, authors indicated that data were available on request, or that all data were presented in the manuscript and its accompanying files, or that data were available from a specified publicly available source (Table 1). Some authors provided information that was not related to the availability of raw data at all; those statements included information related to dissemination of research findings to participants and the public, ethics committee that approved the study protocol, sharing the manuscript content, individuals that collected and analyzed data, dates when the samples were collected, patient informed consent, etc. (Supplementary file 2).

One preprint article declared that data were not publicly available, but that they are available for purchase (Quote: “*We cannot distribute this data, but it is available for purchase to qualified researchers working on projects for the benefit of Medicare beneficiaries.*”).

In 234 (26%) preprints, data/code sharing statement was reported within the full-text manuscript published online on the preprint server. Among 678 articles published on

**Table 1** Categories of data sharing statements reported in the “Data/Code field” on the preprint web site

Category	N (%)
Data available on request	130 (15)
Data generated or used during the study have been presented in the submitted article	96 (11)
Data were posted on a specified website [ <i>name of the website was explicitly reported</i> ] other than GitHub	106 (12)
Data used in the manuscript are available from references, or another [ <i>explicitly reported</i> ] public source	74 (8.2)
Data and script are available in GitHub	55 (6.1)
Data available on <b>reasonable</b> request	41 (4.6)
Authors described a source of their data, but not exact public location	15 (1.7)
Not available	13 (1.4)
Not applicable	13 (1.4)
Statement that was not about data availability	13 (1.4)

Ten most common categories are shown, from the 897 analyzed preprints

medRxiv, 179 (26%) had a data/code sharing statement within the manuscript. Among 219 articles published on bioRxiv, 55 (25%) had a data/code sharing statement within the manuscript; 42 of those 55 did not have a Data/Code field on the article website, and 13 did have it. The ten most common categories of information provided by authors in the data/code sharing statement in the preprint manuscript are shown in Table 2. All other categories can be seen in Supplementary file 2 (column N). Most commonly, authors indicated that data were available on GitHub or another clearly specified web location, on (reasonable) request, in the manuscript or its supplementary files (Table 2).

In 571 (64%) of the analyzed preprint articles, there was a qualitative discrepancy in the information provided in the Data/Code field on the article website and data/code sharing statement within the manuscript.

**Table 2** Ten most commonly used categories of data sharing statements reported in the full text of the preprint manuscript (N = 234)

Category	N (%)
Data and script are available in GitHub	47 (19)
Data available on request	42 (18)
Data were posted on a specified website other than GitHub [ <i>name of the website was explicitly reported</i> ]	38 (16)
Data generated or used during the study have been presented in the submitted article (in the manuscript and/or supplementary materials)	33 (14)
Data available on <b>reasonable</b> request	29 (12)
Data used in the manuscript are available from a public/open source [ <i>but the name of the public source not reported</i> ]	27 (12)
No additional data available	4 (1.7)
Not applicable	4 (1.7)
Data not shared (“Not publicly available”; “Cannot be shared online”; “Data obtained for this study will not be made available to others”)	3 (1.3)
Submitted (“Data are submitted”; “Submitted to databases”)	2 (0.9)

We checked whether those sources indeed contained raw data for 283 (32%) preprints that reported that data were available somewhere (i.e. in the manuscript or online in a repository, etc.). We found raw data/code for 133 (47%) of those 283 preprints. Thus, raw data/code were publicly available for 15% of all the analyzed preprint articles. Among those 133 articles, data were shared by 117 (88%) articles, code by 9 (7%) and both data and code by 7 (5%). In 234 (26%) preprints, authors reported that they obtained their data from elsewhere.

## Publication of preprint articles

By January 2022, 422 (47%) of the included preprint articles were published in a scholarly journal, 474 (53%) were not published, and 1 (0.1%) was withdrawn from the preprint server. The articles were published in 216 different journals. The highest number of articles was published in the following journals: PLoS One (N=18; 4.3%), Science (N=12; 2.8%), Nature (2.6%), Frontiers in Medicine (N=9; 2.1%), International Journal of Infectious Diseases (N=9; 2.1%) and Journal of Medical Virology (N=9; 2.1%).

## Author survey

There were 489 preprints eligible for the author survey, as they belonged to one of the first three categories, regarding their DAS (no DAS, data available on request, raw data not found). We were unable to survey 11 authors because the corresponding author's name or email were not reported in the preprint. Thus, we emailed the survey to 478 authors; one email returned undelivered, and we received 66 (14%) responses from the rest. Among the 66 responders, 25 participants indicated that data were available in their manuscript, but we could not find them. There were 22 participants who indicated that data were available on request, and 19 participants who did not provide any DAS. Their categorized responses are shown in Table 3. Most commonly, the authors explained where the raw data could be found (N=38; 58%). A number of authors responded to our message, but did not provide the response we were looking for, including 8 authors who did not explain what exactly would be a "reasonable request" (Table 3). Several authors explained what would be a reasonable request, including a scientific project, coming from a reputable institution, with a clear aim (Table 3).

## Discussion

This study found that only a quarter of preprint articles on COVID-19, posted on bioRxiv and medRxiv, had a data/code sharing statement within the manuscript. Furthermore, among the preprint articles that reported that data were available somewhere (i.e., in the manuscript or online in a repository, etc.), we found those raw data for less than half of those articles. Overall, 15% of the analyzed preprint articles have publicly shared raw data and/or code.

The results are comparable to the results of Lucas-Dominguez et al., who found that 13.6% of articles retrieved from PubMed Central early in the COVID-19 pandemic made their research data available (Lucas-Dominguez et al., 2021). Even though these data publication rates appear to be low, even lower rates were published for other fields and articles. In 2021, Towse et al. reported that 4% out of 1900 articles from 15

**Table 3** Categorizes responses from author survey on reasons for not sharing their data within the manuscript

Group of authors	Responses	N
<i>Indicated that data were available in the manuscript, but such data were not found (N=25)</i>		
Author explained where the data are available		19
There are no primary data that have been collected as part of the study		2
Did not clarify the location of raw data		2
Not willing to share their data		1
Did not answer the question in the response		1
<i>Indicated in data availability statement that data are available on request (N=22)</i>		
Did not explain what would be a reasonable request for data sharing		8
Author explained where the data are available		6
Reasonable request would be a scientific project or an institutional/healthcare project		1
We share data generated in our study upon request, for instance, upon requests by email		1
Author explained where the data are available / raw data will be published later		1
The data will be published only when the manuscript is formally published		1
Author explained where the data are available / author explained what would be a reasonable request for data sharing		1
Would be motivated to share if the request comes from a reputable institution with a clear aim and objectives for the data analysis		1
Not willing to share their data		1
Did not answer the question in the response		1
<i>The authors did not report data availability statement in the manuscript (N=19)</i>		
Author explained where the data are available		14
Authors will share their raw data but want to know what exactly is planned to do with the data		1
No new data set was used in the paper		1
Not sure what the term raw data means		1
The data will be published only when the manuscript is formally published		1
Willing to share their data but it's difficult concerning the size of data		1

psychological journals have adhered to the open research data (Towse et al., 2021). Gorman analyzed data sharing in 13 high-impact addiction journals and found that only one (0.8%) out of 130 analyzed articles contained a direct link to the analyzed data (Gorman, 2020).

Another issue is the quality and completeness of the shared datasets. Roche et al. analyzed 100 datasets from journals publishing ecological and evolutionary research that have a strong public data archiving policy. They reported that 56% of the analyzed datasets were incomplete, and 64% archived in a manner that partially or entirely prevented their reuse (Roche et al., 2015).

Interest in raw data collected in studies devoted to a public health emergency is not purely academic exercise. During the COVID-19 pandemic, multiple high-profile retractions of research articles have been published; some of them happened when the data analytics company refused to share the raw data. Subsequently, it was suggested that journals should institute mandatory requests to authors to share the primary data as a measure that will likely ensure data integrity and transparency of the research findings and help prevent publication frauds (Krishan & Kanchan, 2020).



While writing a data sharing statement and sharing raw data is not synonymous, our study provides relevant insight into what happens when something is mandatory. Namely, bioRxiv and medRxiv had a different approach to requiring statements regarding data/code availability. The medRxiv requires the following from authors [quote]: “*Please include a statement regarding the availability of all data referred to in the manuscript and note links below.*”, and there is a separate field for Data availability links, indicating [quote] “*Please provide any URLs for external datasets or supplementary material online at other repositories that pertain to this manuscript. These links will be provided online for readers once this submission is posted online. (Example: <https://www.example.com>).*” Since the Data/Code field is obligatory in the medRxiv, this explains why virtually all articles posted on medRxiv had something written in the Data/Code field, compared to less than 10% of articles published on bioRxiv. Obviously, when authors are not required to disclose anything related to their Data/Code, few authors do it voluntarily.

It has been reported that journals could leverage compulsory open data to develop the reputation and amplify their journal impact factor (Zhang & Ma, 2021). While preprint servers are not journals, their obligatory demand for raw research data or code could help amplify their reputation in the field.

The authors should be required to provide their data sharing statement within the manuscript as well, as it is unclear how many readers will look for a Data/Code field on the website of the preprint article. Presumably, readers interested in the study will mostly rely on information provided within the manuscript. Preprint servers bioRxiv and medRxiv should request authors to include data sharing statements within the manuscript as well.

Li et al. analyzed data sharing intentions of COVID-19 clinical trials of interventions, as declared by authors in trial registrations and publications. They included 924 trial registrations in the analysis; authors of 15.7% of registrations were willing to share data, 38.6% were willing to share immediately after publishing results, and 47.6% reported they were unwilling to share their study data. The authors found 28 published COVID-19 clinical trials; of those, only 7 had a data sharing statement, with six that reported authors were willing to share data, and one reported data were not available (Li et al., 2021).

However, we need to be aware that the presence of a data sharing statement and the authors’ self-reported intention to share data may not translate to raw data sharing upon request. We have shown that even authors who indicated in their data availability statement that data will be available on request mostly do not even respond to the data request; few authors of clinical trials were willing to share their data (Gabelica et al., 2019).

Some researchers may need education regarding data sharing issues, as we found multiple statements in the Data/Code field that had nothing to do with data sharing. Despite very clear description about what is expected to be in the Data availability field, some authors wrote strange information in that field, for example, information about competing interests, or information that is difficult to interpret, such as “*All authors agree that all data submitted here are publicly available.*” Furthermore, many authors wrote that “all data” are in the manuscript or accompanying files, but neither the manuscript nor the associated files contained raw data; this implies that authors may not be aware of the meaning of the “data sharing” concept and that data sharing implies sharing of raw data collected within the study.

We even found one case where the authors expect payment for the data (DeCapprio et al., 2020a). Curiously, in the version of the article that was published in a scholarly journal, the authors did not write that a payment is needed to access the data. Instead, the authors simply wrote that the data are proprietary and they are not shareable (DeCapprio et al., 2020b).

Studies such as this one are relevant because they may help reshape biomedicine and biomedical research (Puljak, 2020). Ideas for future studies include repeating the same analysis on published articles about COVID-19. This study focused on preprint articles due to the spike in preprint publications at the beginning of the COVID-19 pandemic (Fidahic et al., 2020). Furthermore, it would be worthwhile to attempt to re-analyze raw data that the authors made available. Due to the heterogeneity of the studies in our sample, we did not attempt to do it; a large team of experts would be needed to attempt re-analysis of data from studies in our sample.

It would be interesting to analyze the inclusion of data/code in non-COVID-19 preprint articles in future studies. We searched the literature, but we were unable to find any such reports for comparison.

A limitation of the study is the low response rate of authors contacted in the survey (14%); this number was limited, but not unsurprising as this was an unsolicited email survey. Furthermore, we did not analyze factors associated with data sharing. It has been shown that some factors, for example, the later career stage of the researchers, are associated with more prevalent data sharing (Dorta-González et al., 2021). We have also analyzed publication rates of the included articles, by January 2022. Almost half of the analyzed articles were published in a scholarly journal by that date. It is possible that perhaps more scholarly articles based on those preprints will be published subsequently.

In conclusion, we found that only a quarter of analyzed preprint articles on COVID-19 included a data sharing statement within their manuscript, and 15% shared their raw data or code publicly, either in the manuscript or elsewhere online, at the time of publication. All preprint servers should require authors to provide data sharing statements that will be included both on the website and in the manuscript. In addition, the education of researchers about the meaning of data sharing would be needed.

**Supplementary Information** The online version contains supplementary material available at <https://doi.org/10.1007/s11192-022-04346-1>.

**Acknowledgements** We are grateful to the individuals who participated in our survey.

**Author contribution** LP conceived the research idea and worked as a project coordinator. JS, AC, TG, RLP, TB and LP were involved in data curation, formal analysis, investigation, methodology, and initial draft writing. All authors revised the manuscript critically for the content.

**Funding** This research received no external funding.

**Data availability** All raw data collected within this study are available in Supplementary file 2 and also published on Open Science Framework (<https://osf.io/6ztec/>).

**Code availability** Not applicable.

## Declarations

**Competing interests** The authors declare no conflict of interest.

**Ethical approval** The study included a survey of corresponding authors of manuscripts deposited in preprint servers. The study protocol was approved by the Ethics Committee of the Catholic University of Croatia (date: April 1, 2020; Klasa: 641-03/20-01/06; Urbroj: 498-03-02-06-02/1-20-02). We began surveying the authors only after we received the approval of the Ethics Committee.

**Consent to participate** All authors gave written consent to participate in the study.


**Consent to publish** Not applicable.

## References

- Besancon, L., Peiffer-Smadja, N., Segalas, C., Jiang, H., Masuzzo, P., Smout, C., Billy, E., Deforet, M., & Leyrat, C. (2021). Open science saves lives: Lessons from the COVID-19 pandemic. *BMC Medical Research Methodology*, 21(1), 117.
- da Silva TJA. (2021). Adjusting the use of preprints to accommodate the “quality” factor in response to COVID-19. *Journal of Taibah University Medical Sciences*, 16(4), 477–481.
- DeCapprio, D., Gartner, J., McCall, C. J., Burgess, T., Kothari, S., & Sayed, S. (2020a). Building a COVID-19 Vulnerability Index. *medRxiv*. <https://doi.org/10.1101/2020.1103.1116.20036723>
- DeCapprio, D., Gartner, J., McCall, C. J., Burgess, T., Kothari, S., & Sayed, S. (2020b). Building a COVID-19 vulnerability index. *Journal of Medical Artificial Intelligence*, 3, 1–10.
- Dorta-González, P., González-Betancor, S. M., & Dorta-González, M. I. (2021). To what extent is researchers’ data-sharing motivated by formal mechanisms of recognition and credit? *Scientometrics*, 126, 2209–2225.
- Federer, L. M., Belter, C. W., Joubert, D. J., Livinski, A., Lu, Y. L., Snyders, L. N., & Thompson, H. (2018). Data sharing in PLOS ONE: An analysis of data availability statements. *PLoS ONE*, 13(5), e0194768.
- Fidahic, M., Nujic, D., Runjic, R., Civljak, M., Markotic, F., Lovric Makaric, Z., & Puljak, L. (2020). Research methodology and characteristics of journal articles with original data, preprint articles and registered clinical trial protocols about COVID-19. *BMC Medical Research Methodology*, 20(1), 161.
- Gabelica, M., Cavar, J., & Puljak, L. (2019). Authors of trials from high-ranking anesthesiology journals were not willing to share raw data. *Journal of Clinical Epidemiology*, 109, 111–116.
- Godlee, F., & Groves, T. (2012). The new BMJ policy on sharing data from drug and device trials. *BMJ*, 345, e7888.
- Gorman, D. M. (2020). Availability of research data in high-impact addiction journals with data sharing policies. *Science and Engineering Ethics*, 26(3), 1625–1632.
- Groves, T. (2010). BMJ policy on data sharing. *BMJ*, 340, c564.
- Homolaks, J., Kodvanj I, Virag D (2020) Preliminary analysis of COVID-19 academic information patterns: a call for open science in the times of closed borders. *Scientometrics* 124, 2687–2701.
- Hoy, M. B. (2020). Rise of the Rxivs: How Preprint Servers are Changing the Publishing Process. *Medical Reference Services Quarterly*, 39(1), 84–89.
- Iacobucci, G. (2019). New preprint server allows earlier sharing of research methods and findings. *BMJ*, 365, 14110.
- Kirkham, J. J., Penfold, N. C., Murphy, F., Boutron, I., Ioannidis, J. P., Polka, J., & Moher, D. (2020). Systematic examination of preprint platforms for use in the medical and biomedical sciences setting. *British Medical Journal Open*, 10(12), e041849.
- Krishan, K., & Kanchan, T. (2020). COVID-19 and the need for stringent rules on data sharing. *Acta Biomedica*, 92(1), e2021174.
- Li, R., von Isenburg, M., Levenstein, M., Neumann, S., Wood, J., & Sim, I. (2021). COVID-19 trials: Declarations of data sharing intentions at trial registration and at publication. *Trials*, 22(1), 153.
- Lucas-Dominguez, R., Alonso-Arroyo, A., Vidal-Infer, A., & Aleixandre-Benavent, R. (2021). The sharing of research data facing the COVID-19 pandemic. *Scientometrics*. <https://doi.org/10.1007/s11192-021-03971-6>
- Modjarrad, K., Moorthy, V. S., Millett, P., Gsell, P. S., Roth, C., & Kieny, M. P. (2016). Developing global norms for sharing data and results during public health emergencies. *PLoS Medicine*, 13(1), e1001935.
- Moorthy, V., Henao Restrepo, A. M., Preziosi, M. P., & Swaminathan, S. (2020). Data sharing for novel coronavirus (COVID-19). *Bulletin of the World Health Organization*, 98(3), 150.
- Poremski, D., Falissard, B., Fegert, J., Witt, A., Ordóñez, A. E., Martin, A., & Fung, D. S. S. (2019). Moving from “personal communication” to “available online at”: Preprint servers enhance the timeliness of scientific exchange. *Child and Adolescent Psychiatry and Mental Health*, 13, 42.
- Pourhoseingholi, M. A., Shojaei, S., & Ashtari, S. (2020). Three potential challenges in studying COVID-19 pandemic data: Chinese statistics, social media, and preprint servers. *Gastroenterol Hepatol Bed Bench*, 13(4), 278–279.
- Puljak, L. (2020). Evidence synthesis and methodological research on evidence in medicine—Why it really is research and it really is medicine. *Journal of Evidence-Based Medicine*, 13(4), 253–254.
- Roche, D. G., Kruuk, L. E., Lanfear, R., & Binning, S. A. (2015). Public data archiving in ecology and evolution: How well are we doing? *PLoS Biology*, 13(11), e1002295.

- Towse, J. N., Ellis, D. A., & Towse, A. S. (2021). Opening Pandora's Box: Peeking inside Psychology's data sharing practices, and seven recommendations for change. *Behavior Research Methods*, 53(4), 1455–1468.
- Watson, M. (2015). When will “open science” become simply “science”? *Genome Biology*. <https://doi.org/10.1186/s13059-015-0669-2>
- Wolkewitz, M., & Puljak, L. (2020). Methodological challenges of analysing COVID-19 data during the pandemic. *BMC Medical Research Methodology*, 20(1), 81.
- Zhang, L., & Ma, L. (2021). Does open data boost journal impact: evidence from Chinese economics. *Scientometrics*. <https://doi.org/10.1007/s11192-021-03897-z>

## Authors and Affiliations

Josip Strcic<sup>1</sup> · Antonia Civljak<sup>2</sup> · Terezija Glozinic<sup>1</sup> · Rafael Leite Pacheco<sup>3</sup> ·  
Tonci Brkovic<sup>4</sup> · Livia Puljak<sup>1</sup> 

Josip Strcic  
josip.strcic@gmail.com

Antonia Civljak  
civljak.antonio@gmail.com

Terezija Glozinic  
teraglo33@gmail.com

Rafael Leite Pacheco  
rleitpacheco@hotmail.com

Tonci Brkovic  
toncibrkovic@gmail.com

<sup>1</sup> Center for Evidence-Based Medicine and Health Care, Catholic University of Croatia, Ilica 242, 10000 Zagreb, Croatia

<sup>2</sup> Specialist Family Medicine Clinic Dr. Ljiljana Lipovac-Francuz, Jurkovicева 11, 10000 Zagreb, Croatia

<sup>3</sup> Hospital Sírio-Libanês, Universidade Federal de São Paulo (Unifesp) and Centro Universitário São Camilo (CUSC), São Paulo, Brazil

<sup>4</sup> Department of Nephrology, University Hospital Split, Split, Croatia