

Transparency of COVID-19-related research in medical journals

Ahmad Sofi-Mahmudi^{*†} Eero Raittio^{*‡} Sergio E. Uribe^{§¶}

Corresponding author: Eero Raittio; **Address:** University of Eastern Finland, Institute of Dentistry, P.O. Box 1627, FI-70211 Kuopio, Finland; **Telephone:** +358 294 45 1111; **Email:** eero.raittio@uef.fi.

Authors' contributions: All authors conceived and designed, and wrote and revised the protocol.

What is new

- key findings
- what this adds to what is known
- what is the implication/what should change now

Before completing your submission, please take some time to check that your submission includes “What is new” text, including up to 5 points that address the following areas: key findings, what this adds to what is known, and what is the implication/what should change now.

^{*}Seqiz Health Network, Kurdistan University of Medical Sciences, Seqiz, Iran.; a.sofimahmudi@gmail.com

[†]Cochrane Iran Associate Centre, National Institute for Medical Research Development (NIMAD), Tehran, Iran.

[‡]Institute of Dentistry, University of Eastern Finland, Kuopio, Finland.; eero.raittio@uef.fi

[§]Department of Conservative Dentistry and Oral Health & Bioinformatics Research Unit, Riga Stradins University, Riga, Latvia.; sergio.uribe@rsu.lv

[¶]School of Dentistry, Universidad Austral de Chile, Valdivia, Chile

Abstract

Objective: We aimed to assess the adherence to five transparent practices (data availability, code availability, statements of protocol registration and conflicts of interest and funding disclosures) from open access Coronavirus disease 2019 (COVID-19) related articles.

Methods: We searched and exported all open-access COVID-19-related articles from PubMed-indexed journals available in the Europe PubMed Central database published between January 2020 and December 2021. We detected transparent practices with a validated and automated tool. Basic journal- and article-related information were retrieved from the database. We used R for the searches (europepmc package), data handling, analysis (rtransparent package) and reporting (ggplot2 package).

Results: The total number of COVID-19-related articles was 175,550, of which we were able to retrieve full texts of 132,598 (76%) articles from the database. Of those, 87.0% (95% CI: 86.8%–87.2%) provided conflicts of interest disclosure and 68.9% (95% CI: 68.7%–69.2%) funding disclosure and 3.6% (95% CI: 3.5%–3.7%) had registered protocol. 8.7% (95% CI: 8.6%–8.9%) of studies shared raw data and 3.1% (95% CI: 3.0%–3.2%) of studies shared code. Transparent practices were more common in articles published in 2022 than in 2021 and 2020. Apart from COI disclosure, articles that adhered to transparency practices were published in journals with slightly higher median Journal Impact Factor but received equal citations.

Conclusion: While the majority of the articles were freely available and had a COI disclosure, the adherence to other transparent practices was far from the acceptable level. The adherence was even surprisingly low, particularly in terms of data sharing, taking into account global initiatives to improve open science in COVID-19 related research. A much stronger commitment to open science practices, particularly to protocol registration, data and code sharing, is needed from all stakeholders.

Keywords: COVID-19; Data; FAIR data principles; Open Science.

Introduction

During the recent COVID-19 pandemic, part of the public in developed countries justified their lack of adherence to public health measures due to the lack of transparency of the scientific studies used as an argument by national authorities (1). Thus, supranational bodies such as the OECD indicate that transparent communication to citizens is vital to support public health measures and to counter misinformation (2). On the other hand, the COVID-19 pandemic launched an incredibly fast and voluminous avalanche of scientific publications across the different fields of science. This research led to the development of vaccines as well as the evaluation of community interventions. On the one hand, it is possible that transparent scientific practices, like data and code sharing, have helped to fight pandemic globally (3,4), while on the other hand, the lack of transparency of some key developments has led the public

to mistrust the results of research and public health measures (1). This shows that there is a lack of awareness about the level of transparency practices in COVID-19-related medical literature. As an example, it is unknown whether open science initiatives related to COVID-19 pandemic (3,4) to advance data sharing, has led to higher number of articles sharing data in COVID-19 research compared to what has been seen in studies in general in biomedical literature (5).

Our aim was to assess the adherence to transparent scientific practices (data sharing, code sharing, conflict of interest, COI, disclosures, funding disclosures, and protocol registration) from open access full-text COVID-19-related articles published in PubMed indexed journals from the Europe PubMed Central (EPMC) database. We mapped adherence to transparent practices across time, journal impact factor (JIF) and citations articles have received

Methods

The protocol of this study was published beforehand on the Open Science Framework (OSF) website (osf.io/5kx2n). All code and data related to the study were shared via its OSF repository (osf.io/x3kb6) at the time of submission of the manuscript.

Data sources and study selection

We searched for open access COVID-19-related PubMed-indexed articles available in the EPMC database. We restricted our search to papers in English and considered papers published from 2020-01-01 to 2021-12-31 with variants of COVID-19 keyword in their title, keywords, or results. This enabled us to bypass screening for COVID-19-related papers and decrease the human labor for our research. We confirmed our approach by choosing a random sample of 500 papers and of those, all but 10 were relevant to COVID-19.

The search query was as follows:

```
(COVID-19 query for title, keywords, and results) AND (SRC:"MED") AND (LANG:"eng" OR LANG:"en" OR LANG:"us") AND (FIRST_PDATE:[2020-01-01 TO 2021-12-31]) AND (OPEN_ACCESS:y) AND (PUB_TYPE:"Journal Article" OR PUB_TYPE:"research-article" OR PUB_TYPE:"rapid-communication" OR PUB_TYPE:"product-review")'
```

The full COVID-19 query is available in Appendix 1.

We downloaded all identified available records in XML full-text format (for full-text evaluation) using the `metareadr` package (6) from the database.

Data extraction and synthesis

We assessed articles' adherence to five transparent practices:

- data sharing,
- code sharing,
- COI disclosures,
- funding disclosures,
- protocol registration.

To do so, we used a validated and automated tool developed by Serghiou et al. (5) suitable to identify these five transparent practices from articles in XML format from the EPMC database. Regarding data sharing, code sharing and protocol registration the tool identifies whether the article had actually adhered to practice (for instance “data available on request” are not counted as sharing data), whereas COI and funding disclosure merely detect whether articles include the disclosure or not (regardless of its specific content) (5).

Basic journal- and article-related information (publisher, publication year, citations to article and journal name) were retrieved from the EPMC database. We downloaded additional information about journal impact factors from the Journal Citation Reports (jcr.clarivate.com).

We also calculated the proportion of articles available as open-access full-texts via the Europe PMC from the total number of COVID-19-related articles in the database.

Data analysis

We used R v4.1.2 (7) for searches, data handling, analysis and reporting. The searches and data export from the Europe PMC were conducted with the `europemc` package (8). Indicators of transparency practices from the available full-texts were extracted with the `rtransparent` package (9). Trends over time in transparency practices were reported using descriptive tabulations and graphical illustrations using the `ggplot2` package (10). We used the sensitivity and specificity of the `rtransparent` package (5) to generate 95% confidence intervals (CIs) for our prevalence estimates of the transparency practices with the `epiR` package (11).

Additionally, we calculated how many transparency practices articles had adhered to (from 0 to 5 practices). We used Chi-square for comparing indicators between years and the Wilcoxon rank sum test to test whether there was a relationship between transparency indicators and JIF or received citations. Fisher exact tests with Monte-Carlo simulated p-values for differences in transparency practices by journals and publishers were performed. The thorough information, data and code for the analysis are available as supplementary material on the OSF repository (osf.io/x3kb6).

Results

The total number of COVID-19-related articles (open access and non-open access) was 175,544, of which 133,467 (76.0%) were accessible as open-access in the EPMC. However, 869 (0.7%) were not downloadable because of technical issues and thus were excluded from our analyses. Of those 132,598 remaining articles, 47,548 were published in 2020, 75,842 in 2021, and 9,208 in the first two months of 2022.

These articles were published in 6,020 journals of which the top five were International Journal of Environmental Research and Public Health ($n=3,710$), PLoS ONE ($n=3,059$), Scientific Reports ($n=1,716$), Frontiers in Public Health ($n=1,351$), and Frontiers in Psychology ($n=1,341$). The mean and the median of the number of citations to these articles were 12.0 (standard deviation, $SD = 92.28$) and 2 (interquartile range, $IQR = 7$). The highest cited article with 15,549 citations was an article published in the Lancet in 2020 (12).

About nine-tenth of the articles ($n=115,387$) had a statement to disclose COI (87.0%, 95% CI: 86.8%–87.2%). Articles published in 2022 had a higher rate of COI disclosure (91.4% (2022), 89.1% (2021), 82.9% (2020), $P<0.001$). More than two-thirds of the references had a funding statement ($n=91,413$) (68.9%, 95% CI: 68.7%–69.2%). Articles in 2022 had a funding statement more frequently than two other years (79.0% (2022), 73.1% (2021), 60.4% (2020), $P<0.001$). Less than 1 in 25 of the articles ($n=4708$) were registered beforehand (3.6%, 95% CI: 3.5%–3.7%) and the rate was higher in 2022 (5.4% (2022), 4.2% (2021), 2.2% (2020), $P<0.001$).

Regarding data and code availability, less than 1 in 11 articles had shared their data ($n=11,551$) (8.7%, 95% CI: 8.6%–8.9%). Articles in 2022 had shared data more frequently (11.9% (2022), 9.6% (2021), 6.7% (2020), $P<0.001$). Less than 1 in 30 of the articles ($n=4,055$) had shared their code (3.1%, 95% CI: 3.0%–3.2%) and the rate was higher in 2022 (4.0% (2022), 3.5% (2021), 2.2 (2020), $P<0.001$).

There was a consistently increasing monthly trend in the proportion of articles with COI disclosure, funding disclosure and protocol registration, and the proportions were the highest in February 2022 (Figure 1B). Increases in data and code sharing were weaker than in other transparency practices over time.

One-eleventh of the articles (9.0%, $n=11,889$) did not adhere to any of the five transparency practices, 24.0% ($n=31,836$) adhered to one, 56.0% ($n=74,075$) adhered to two, and 9.1% ($n=12,109$) adhered to three. Few (2.0%, $n=2,644$) articles showed adherence to four transparency practices, and less than 0.1% of the articles ($n=45$) had used all five practices.

There were journal and publisher-related statistically significant differences in all five practices ($p<0.001$ for all), except for code sharing based on JIF (0.14) (Table 1 and 2, Supplementary material).

Apart from COI disclosure, articles that adhered to transparency practices were published in journals with slightly higher median JIF but received equal citations (Table 3).

Discussion

Adherence to transparent practices increased over time in COVID-19 related medical literature during 2020-2021. Adherence to reporting COI disclosure was high throughout study period, in addition most of articles had funding disclosure. Data sharing, code sharing and protocol registration were rare, but improved little over study period. COI disclosures seemed to be more common in articles published in lower impact journals, otherwise differences in transparent practices according to received citations and JIF were small.

Adherence to transparent practices was at similar level in COVID-19 related literature than in other biomedical literature analyzed with the same methods (see Fig 3 in (5)). This is surprising particularly when considering worldwide, remarkable and noble initiatives related to enhancing open science to tackle the pandemic. As early as in January 2020, over 100 organizations, including journals, publishers, funders, universities and other institutions, signed a statement that aimed to ensure free access to research data, tools and other information related to COVID-19 (4,13). Later also other initiatives to support the goal emerged, for instance COVID-19 Open Research Dataset (CORD-19)(14), a free resource of over 280,000 articles about the COVID-19 virus. However, it is possible that the algorithms used here did not efficiently detect all different ways of sharing data and material emerged after the pandemic, because the algorithms were validated before the pandemic (5).

In comparison to other studies on the transparency of COVID-19 related research is difficult due to different methodologies and scale of our research. However, it seems that data sharing was little less common than in our sample than what was the proportion detected in the beginning of the pandemic measured by the investigation of associated data in PubMed Central (15). In addition, the proportion of studies which shared data in our study sample was lower than in COVID-19 related preprints shared via medRxiv and bioRxiv (16). Lacking protocol registration, sharing code and COI have got attention during the pandemic (3,17,18), however we are not aware of the investigation of transparency in these aspects in COVID-19 related research in this scale in medical research.

Despite we used representative data and validated methods to investigate transparency in COVID-19-related research, our study has some weaknesses. The study sample was restricted to open access articles in the EPMC database which may not correspond to all COVID-19-related studies published in medical journals. At least in general, the differences in transparency practices between open access and non-open access articles from Europe PMC are rather similar (19). However, as the majority of COVID-19-related papers were available via EPMC (76%), this does not diminish the strength of our interpretations considerably.

Transparency is crucial to ensure the credibility of science and enable its assessment. Transparent scientific practices, like the ones we investigated here, have helped to fight pandemic globally (3,4). Whilst majority of COVID-19-related articles were open access and had adhered

to disclose funding and COI, your findings showed suboptimal adherence to data and code sharing and protocol registration. Stronger commitment to open science practices, particularly to protocol registration, data and code sharing, is needed from all stakeholders.

Acknowledgements

The computational analyzes were performed on servers provided by UEF Bioinformatics Center, University of Eastern Finland, Finland. The study was supported by a personal grant to ER from the Finnish Dental Society Apollonia.

References

Tables

Table 1. Transparency practices in the six most common journals in the sample.

| | BMC Oral Health, N = 2508 | Dental Press J Or- thod, N = 530 | Head Face Med, N = 572 | J Appl Oral Sci, N = 1440 | J Indian Prosthodont Soc, N = 495 | Med Oral Patol Oral Cir- dental, N = 1174 | Other, N = 3945 | p- value |
|-------------------------------|---------------------------------------|---|------------------------------------|------------------------------------|---|--|-----------------------|-------------|
| COI dis- closure | 2489 (99%) | 502 (95%) | 537 (94%) | 303 (21%) | 301 (61%) | 869 (74%) | 3172 (80%) | <0.001 |
| Funding disclo- sure | 2232 (89%) | 87 (16%) | 280 (49%) | 656 (46%) | 301 (61%) | 497 (42%) | 2505 (63%) | <0.001 |
| Protocol registra- tion | 326 (13%) | 12 (2.3%) | 23 (4.0%) | 31 (2.2%) | 4 (0.8%) | 43 (3.7%) | 298 (7.6%) | <0.001 |
| Data sharing | 120 (4.8%) | 2 (0.4%) | 2 (0.3%) | 7 (0.5%) | 0 (0%) | 5 (0.4%) | 76 (1.9%) | <0.001 |
| Code sharing | 6 (0.2%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0 (0%) | 0.61 |

P-values from Fisher's Exact Test for Count Data with simulated p-value (based on 2000

replicates). COI: Conflict of interest.

Table 2. Transparency practices in the six most common publishers of journals in the sample.

| Measurement | BioMed Central, N = 3075 | Dental Press Editora, N = 530 | Faculdade Medicina de Odon- tologia de Bauru, N = 1440 | Oral, Patolo- gia Oral y Cirugia Bucal, N = 1174 | Springer Verlag, N = 935 | Wolters Kluwer Med- know Publica- tions, N = 495 | Other, N = 2619 | p-value |
|-------------------------------|-----------------------------|----------------------------------|--|--|-----------------------------|--|--------------------|---------|
| | | | | | | | | |
| COI dis- closure | 3026 (98%) | 502 (95%) | 303 (21%) | 869 (74%) | 904 (97%) | 301 (61%) | 1878 (72%) | <0.001 |
| Funding disclo- sure | 2512 (82%) | 87 (16%) | 656 (46%) | 497 (42%) | 596 (64%) | 301 (61%) | 1593 (61%) | <0.001 |
| Protocol registra- tion | 349 (11%) | 12 (2.3%) | 31 (2.2%) | 43 (3.7%) | 113 (12%) | 4 (0.8%) | 144 (5.5%) | <0.001 |
| Data sharing | 122 (4.0%) | 2 (0.4%) | 7 (0.5%) | 5 (0.4%) | 13 (1.4%) | 0 (0%) | 54 (2.1%) | <0.001 |
| Code sharing | 6 (0.2%) | 0 (0%) | 0 (0%) | 0 (0%) | 1 (0.1%) | 0 (0%) | 1 (<0.1%) | 0.68 |

P-values from Fisher's Exact Test for Count Data with simulated p-value (based on 2000 replicates). 391 articles were published in journal without available publisher information. COI: Conflict of interest.

Table 3. Citations to article and journal impact factor by transparency practices.

| Measurement | Citations to article | | P-Value | Journal Impact Factor | | P-value |
|-----------------------|-------------------------|---------|---------|-----------------------------|------------|---------|
| | Median (IQR) | | | Median (IQR) | | |
| | With | Without | | With | Without | |
| COI disclosure | 2 (5) | 5 (9) | < 0.001 | 2.8 (0.61) | 2.7 (0.06) | < 0.001 |
| Funding disclosure | 2 (5) | 3 (7) | < 0.001 | 2.8 (0.06) | 2.7 (0.70) | < 0.001 |
| Registration | 1 (4) | 2 (7) | < 0.001 | 2.8 (0.01) | 2.7 (0.61) | < 0.001 |

| | Citations to article | | | Journal Impact Factor | | |
|-----------------|-------------------------|-------|------|-----------------------------|------------|---------|
| Data sharing | 2 (4) | 2 (7) | 0.01 | 2.8 (0.00) | 2.8 (0.49) | < 0.001 |
| Code sharing | 1.5 (3) | 2 (7) | 0.70 | 2.8 (0.00) | 2.8 (0.49) | 0.08 |

P-value based on the Wilcoxon rank sum test. 1929 articles were published in journals with no impact factor. COI: Conflict of interest; SD: Standard deviation; IQR: Inter-quartile range.

1. Skipper M. COVID vaccine confidence requires radical transparency. Nature [Internet]. 2020 Oct;586(7827):8. Available from: <http://dx.doi.org/10.1038/d41586-020-02738-y>
2. Bellantoni A, Badr K, Alfonsi C. Transparency, communication and trust: The role of public communication in responding to the wave of disinformation about the new coronavirus [Internet]. <https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/>; Available from: <https://www.oecd.org/coronavirus/policy-responses/transparency-communication-and-trust-the-role-of-public-communication-in-responding-to-the-wave-of-disinformation-about-the-new-coronavirus-bef7ad6e/>
3. Besançon L, Peiffer-Smadja N, Segalas C, Jiang H, Masuzzo P, Smout C, et al. Open science saves lives: lessons from the COVID-19 pandemic. BMC Medical Research Methodology [Internet]. 2021 Jun 5;21(1). Available from: <http://dx.doi.org/10.1186/s12874-021-01304-y>
4. Vuong Q-H, Le T-T, La V-P, Nguyen HTT, Ho M-T, Van Khuc Q, et al. Covid-19 vaccines production and societal immunization under the serendipity-mindsponge-3D knowledge management theory and conceptual framework. Humanities and Social Sciences Communications [Internet]. 2022 Jan 18;9(1). Available from: <http://dx.doi.org/10.1057/s41599-022-01034-6>
5. Serghiou S, Contopoulos-Ioannidis DG, Boyack KW, Riedel N, Wallach JD, Ioannidis JPA. Assessment of transparency indicators across the biomedical literature: How open is open? Bero L, editor. PLOS Biology [Internet]. 2021 Mar 1;19(3):e3001107. Available from: <http://dx.doi.org/10.1371/journal.pbio.3001107>
6. Serghiou S. Metareadr: Downloads data often needed for meta-research. 2022; Available from: <https://github.com/serghiou/metareadr>
7. R Core Team. R: A language and environment for statistical computing. 2022; Available from: <https://www.R-project.org/>
8. Jahn N. Europepmc: R interface to the europe PubMed central RESTful web service. 2021; Available from: <https://CRAN.R-project.org/package=europepmc>

9. Serghiou S. Rtransparent: Identifies indicators of transparency. 2021; Available from: <http://github.com/serghiou/rtransparent>
10. Wickham H. ggplot2: Elegant graphics for data analysis. 2016; Available from: <https://ggplot2.tidyverse.org>
11. Stevenson M, Nunes ES with contributions from T, Heuer C, Marshall J, Sanchez J, Thornton R, et al. epiR: Tools for the analysis of epidemiological data. 2022; Available from: <https://CRAN.R-project.org/package=epiR>
12. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, et al. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. The Lancet [Internet]. 2020 Feb;395(10223):497–506. Available from: [http://dx.doi.org/10.1016/s0140-6736\(20\)30183-5](http://dx.doi.org/10.1016/s0140-6736(20)30183-5)
13. Coronavirus (COVID-19): sharing research data [Internet]. Available from: <https://wellcome.org/press-release/sharing-research-data-and-findings-relevant-novel-coronavirus-ncov-outbreak>
14. CORD-19 | Semantic Scholar [Internet]. Available from: https://www.semanticscholar.org/cord19?utm_campaign=CORD-19&utm_source=hs_email&utm_medium=email&utm_content=85059182&_hsenc=p2ANqtz-_AzPEfvedG_BmLwzXX7INzZ88WRx2bF75yj1q7eG4VbJ1RjtQFBfC0z4JvHJo31krm8BKQjkkx0WlPI2R2blZ-GyO2MCQ&_hsmi=85059182
15. Lucas-Dominguez R, Alonso-Arroyo A, Vidal-Infer A, Aleixandre-Benavent R. The sharing of research data facing the COVID-19 pandemic. Scientometrics [Internet]. 2021 Apr 26;126(6):4975–90. Available from: <http://dx.doi.org/10.1007/s11192-021-03971-6>
16. Strcic J, Civljak A, Glozinic 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 [Internet]. 2022 Mar 25; Available from: <http://dx.doi.org/10.1007/s11192-022-04346-1>
- 17.
18. Luengo-Oroz M, Hoffmann Pham K, Bullock J, Kirkpatrick R, Luccioni A, Rubel S, et al. Artificial intelligence cooperation to support the global response to COVID-19. Nature Machine Intelligence [Internet]. 2020 May 22;2(6):295–7. Available from: <http://dx.doi.org/10.1038/s42256-020-0184-3>
19. Wallach JD, Boyack KW, Ioannidis JPA. Reproducible research practices, transparency, and open access data in the biomedical literature, 2015–2017. Dirnagl U, editor. PLOS Biology [Internet]. 2018 Nov 20;16(11):e2006930. Available from: <http://dx.doi.org/10.1371/journal.pbio.2006930>