# Front Matter

**Title:**   
All I want for Christmas is you(r data): The impact of journal data-sharing policies on data availability statements

**Authors and Affiliations**

Luke A McGuinness1,2 (ORCID: 0000-0001-8730-9761), Athena L Sheppard3 (ORCID: 0000-0003-1564-0740)

1. MRC Integrative Epidemiology Unit at the University of Bristol, Bristol, UK
2. Population Health Sciences, Bristol Medical School, University of Bristol, Bristol, UK
3. Department of Health Sciences, University of Leicester, Leicester, UK

**Corresponding author:**

Luke McGuinness; Bristol Medical School, University of Bristol, Canynge Hall, 39 Whatley Road, Bristol, BS8 2PS, United Kingdom; [luke.mcguinness@bristol.ac.uk](mailto:luke.mcguinness@bristol.ac.uk)

**Keywords**

Preprints; Observational study; Reproducibility; data sharing

##### 

# Abstract

**Objective**  
To assess the distribution of preprinted data availability statements across varying categories of “openness”, and to assess whether strict journal data-sharing policies result in a change in “openness” for preprints that are subsequently published.

**Design**  
Observational study of the data availability statements accompanying preprints posted on the medRxiv repository between 25th June 2019 and 1st May 2020, and their published counterparts.

**Setting**  
medRxiv preprint repository.

**Results**  
A classification system was developed to categorize data availability statements as ‘open’ or ‘closed’. Data availability statements from 4101 preprints posted in medRxiv were categorized by two independent researchers. 159 (3.9%) of preprints had subsequently been published and contained evaluable data availability statements. 911 (22.2%) of the preprints had made their data open, 3027 (73.8%) made their data closed and for 163 (4.0%) preprints data availability was not applicable (e.g. editorial, protocol for a systematic review). Of the 159 published articles, 59 (37.1%) made their data open, 96 (60.4%) made their data closed and for 4 (2.5%) articles data availability was not applicable. Of the 151 records for which the article was published and neither the preprint nor published article were categorized as not applicable, 57 (37.7%) were published in journals which mandated open data sharing. Data more frequently became open between preprinting and publication when the journal mandated open data sharing (open at preprint: 33.3% (19), open at publication: 61.4% (35)) compared to when the journal did not mandate open data sharing (open at preprint: 20.2% (19), open at publication: 22.3% (21)).

**Conclusion**  
Requiring that authors submit a data availability statement, is a good first step, but certainly not the last in ensuring transparency of data availability reporting. Strict editorial policies that require data sharing (where appropriate) as a condition of publication appear to be effective in making research data available. We recommend that journals adopt these stricter data-sharing policies which, along with a dedicated code availability statement, will aid in increasing the reproducibility of published results.

##### 

# Introduction

## Background (3 paragraphs)

**1. Brief introduction to the area – what is a DAS and why are they important.**

Sharing research data is becoming an increasingly important aspect of scientific research. Without data sharing, researchers are unable to examine, verify or build on the results of a study, leading to the same data being generated for multiple studies, stifling secondary research and preventing researchers from verifying each other’s work.

As a result, many journals now require data availability statements. These are dedicated sections of research articles, which are intended to provide readers with important information about whether the data described by the study are available and if so, where they can be obtained.[2]

While requiring data availability statements is an admirable first step for journals to take, a lack of regulation on what can be included in these statements often leads to issues. Many authors claim that their data can be made available on request, which may seem like a reasonable approach to sharing research data, however, previous work has shown that there are many flaws with this method of data sharing. When data is requested, it is not actually made available in the majority of cases.[3,4

**2. Description of the problem – while a DAS is often a journal requirement, . It was clear from looking at the journal policies that almost all are keen for authors to share their data but a) authors are doing the bare minimum, and b) journals are not critically reviewing DAS or enforcing their policies harshly enough.**

This suggests that requiring data availability statements without a corresponding, strictly enforced data-sharing policy, these statements are not achieving their aim of making research data more openly available. However few journals require data sharing as a condition of publication. Of a sample of 318 biomedical journals, only ~20% had a data-sharing policy that required data sharing.[6]

**3. Current evidence base – other studies have looked at DAS in published articles but no one has looked at preprints to date. This is especially relevant given the role of preprints in the response to the ongoing COVID-19 pandemic. This leads nicely into our aims/objectives.**

Several previous studies have examined the data availability statements of published articles,[2,7–9] but to date, none has examined the statements accompanying preprinted manuscripts, including those hosted on medRxiv, the preprint repository for manuscripts in the medical, clinical, and related health sciences.[10]

Given that preprints, particularly hosted on medRxiv, have impacted the academic discourse around the recent (and ongoing) COVID-19 pandemic to a similar, if not greater, extent than published manuscripts,[11] assessing the “openness” of their data availability statements is worthwhile.

By comparing the preprint and published data availability statements for the same paper, the potential impact of journal data-sharing policies can be examined.

\_\_Aims\_

This study aimed to explore the distribution of data availability statements across the a number of categories – as listed in Table 1 - and to assess the change between preprint and published data availability statements, stratified by journal data-sharing policy. We also intended to examine whether authors planning to make the data available on publication actually do so, and whether data availability statements are sufficient to capture code availability declarations.

##### 

# Methods

## Protocol, ethics and reporting

A protocol for this analysis was registered in advance and followed at all stages of the study.[12] Any deviations from the protocol are described. Ethical approval was not required for this study.

## Data extraction

The data availability statements of preprints posted on the medRxiv preprint repository between 25th June 2019 (the date of first publication of a preprint on medRxiv) and 1st May 2020 were extracted using the medrxivr and rvest R packages.[13,14] Details on the journal in which preprints were subsequently published was extracted using the published DOI provided by medRxiv and rcrossref.[15] Several other R packages were used for data cleaning and analysis. [16–21]

The data availability statements for published articles were extracted manually into an Excel file, and are available for inspection (see Material availability section).

## Analysis

A classification system was developed to categorize each data availability statement as either open or closed, with additional ordered sub-categories indicating the degree of openness (see Table 1). The system was based on the Findability and Accessibility elements of the FAIR framework,[**???**] the categories used by previous effort to categories published data availability statements,[2,7] and discussion with colleagues. The data availability statement for each preprinted record were categorized by two independent researchers, using the groups presented in Table 1, while the statements for published articles were categorized using all groups barring Category 3 and 4 (“Available in the future”). Researchers were provided only with the data availability statement, and as a result, were blind to the associated preprint metadata (e.g. title, authors, corresponding author institution) in case this could affect their assessments. Any disagreements were resolved through discussion. Due to our large sample, if authors claimed that all data were available in the manuscript or as a supplemental file, or that their study did not make use of any data, we took them at their word. Where a data availability statement met multiple categories, or contained multiple data sources with varying levels of openness, we took a conservative approach and categorized it on the basis of the most restrictive aspect (see Supplementary Materials 3 for some illustrative examples). We plotted the distribution of preprint and published data availability statements across the nine categories presented in Table 1. Records for which the data availability statement was categorized as “Not applicable” (Category 1 from Table 1) at either the preprint or published stage were excluded from further analyses.

Table 1: Categories used to classify the data availability statements

| **Key** | **Main category** | **Sub-category** | **Example** |
| --- | --- | --- | --- |
| **0** | Not applicable (protocol for a review, commentary, etc) |  | "Data sharing not applicable to this article as no datasets were generated or analysed during the current study."a |
| **1** | "Closed" | Data not made available | "Not available for public"b |
| **2** | "Closed" | Data available on request to authors | "Data can be available upon reasonable request to the corresponding author."c |
| **3** | "Closed" | Data will be made available in the future (link provided) | "The protocol and full dataset will be available at Open Science Framework upon peer review publication (https://osf.io/rvbuy/)."d |
| **4** | "Closed" | Data will be made available in the future (no link provided) | "Data will be deposited in Dryad upon publication"e |
| **5** | "Closed" | Data available from central repository (access-controlled or open access), but insufficient detail available to find specific dataset | "Data were obtained from the international MSBase cohort study. Information regarding data availability can be obtained at https://www.msbase.org/." OR  Daily diagnosis number of countries outside China is download from WHO situation reports (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports). https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reportsf |
| **6** | "Closed" | Data available from central access-controlled repository, and sufficient details included to identify specific dataset e.g. via extract or accession ID or date stamp | "This research has been conducted using the UK Biobank Resource under application number 24494. All bona fide researchers can apply to use the UK Biobank resource for health related research that is in the public interest."g |
| **7** | "Open" | Data available in the manuscript/supplementary files | "All data related to this study are present in the paper or the Supplementary Materials. . ."h |
| **8** | "Open" | Data available via a online repository that is not access-controlled e.g. GitHub, Zenodo | "Extracted data used in this meta-analysis and analysis code are available at www.doi.org/10.5281/zenodo.3149365."i |
| a[@ehrlich2019]; b[@septiandri2019]; c[@solis2019]; d[@ebbeling2019]; e[10.1101/2020.04.08.20057927]; f[@malpas2019]; g[@knuppel2019]; h[@thompson2019]; i[@moriarty2019a] | | | |

To assess if data availability statements change between preprint and published articles, we examined whether a discrepancy existed between the categories assigned to the preprinted and published statements, and the direction of the discrepancy (more “closed” or more “open”). We declare a minor deviation from our protocol,[12] in relation to this analysis. Rather than investigating the data-sharing policy only for journals with the greatest change in openness, we extracted and categorized the data-sharing policies for all journals for which preprints had subsequently been published using two categories (1: “requiring/mandating data sharing” and, 2: “not requiring/mandating data sharing”), and compared the change in openness between these two categories.

To assess claims that data will be provided on publication, the data availability statements accompanying the published articles for all records in Category 3 (“Data available on publication (link provided)”) or Category 4 (“Data available on publication (no link provided)”) from Table 1 were assessed, and any difference between the two categories examined. Finally, to assess whether data availability statements also capture code availability, the data availability statement and full text PDF for a random sample 400 preprinted records were assessed for code availability (1: “code availability described” and 2: “code availability not described”).

## Patient and public involvement

Due to the study design and topic, patients and the public were not involved in the choice of research question, the design the study, the conduct of the study, the interpretation of the results, or our dissemination plans. Dissemination to participants is not applicable.

##### 

# Results

The data availability statements accompanying 4101 preprints - registered between 25th June 2019 and 1st May 2020 - were extracted from the medRxiv preprint repository on the 26th May 2020 and were coded according to the categories in Table 1. During the coding process, agreement between raters was high (Cohen’s Kappa = 0.98; “almost perfect agreement”).

Of the 4101 preprints, 163 (4.0%) in Category 0 (“Not applicable”) were excluded following coding, leaving 3938 remaining records. Of these, 911 (23.1%) had made their data open as per the criteria in Table 1. The distribution of data availability statements across the categories can be seen in Figure 1. A total of 379 (9.2%) preprints had been subsequently published, and of these, only 159 (42.0%) had data availability statements that we could categorize. 4 (2.5%) records in Category 0 (“Not applicable”) were excluded, and of the 155 remaining, 59 (38.1%) had made their data open as per our criteria.

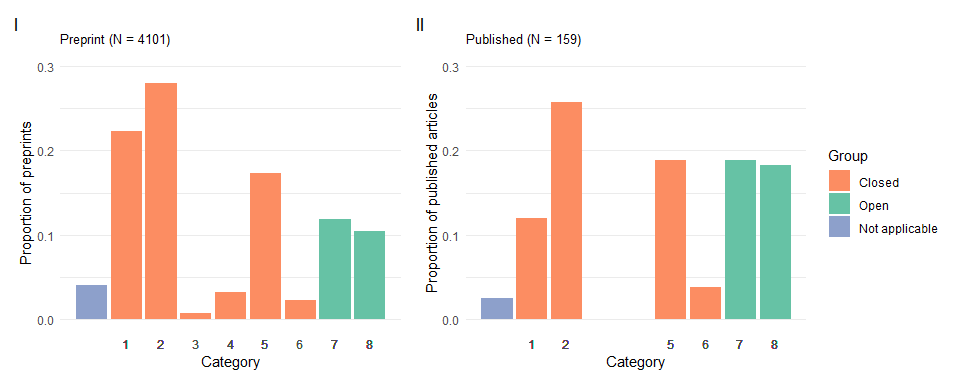


Figure 1: Distribution of the data availability statements of preprinted and published records by category. The categories from Table 1 are presented on the X-axis.

For the comparison of preprint data availability statements with their published counterparts, we excluded records that were not published, that did not have a published data availability statement or that were labeled as “Not applicable” at either the preprint or published stage, leaving 151 records (3.7% of the total sample of 4101 records) records. When grouped by data-sharing policy, there was a greater change towards open data availability statements in journals requiring/mandating data sharing versus those that encouraged it (Table 2). Moreover, a number of articles published in journals that did not require open data sharing, became less open on publication. Change in openness for each individual journal is shown in Supplementary Table 1, while the change for preprints grouped by category, and stratified by journal policy, is shown in Supplementary Table 2.

Table 2: Change in openness of data availability statements from preprint to published arcticle, grouped by journal data-sharing policy.

| **Policy category** | **Number of journals  (N)** | **Number of records  (N)** | **Open at preprint  % (N)** | **Open at publication  % (N)** | **Change from preprint to publication** | | |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **More open  (N)** | **More closed  (N)** | **No change  (N)** |
| **Does not require open data** | 70 | 94 | 20.2% (19) | 22.3% (21) | 10 | 8 | 76 |
| **Requires open data** | 20 | 57 | 33.3% (19) | 61.4% (35) | 16 | 0 | 41 |

161 (3.9%) preprints stated that data would be available on publication, but only 10 of these had subsequently been published (Table 3) and openness on publication did not seem to vary based on whether the preprinted data availability statements include a link to an embargoed repository or not (though the sample size is small).

Table 3: Assessment of whether researchers promising to make data available on publication actually do so, and whether this differs if researchers included a link to an embargoed repository or not.

| **Group** | **Number of records** | **Open on publication** |
| --- | --- | --- |
| **Available in future (link)** | 3 | 1 (33.33%) |
| **Available in future (no link)** | 7 | 5 (71.43%) |

Of the 400 records for which code availability was assessed, 75 mentioned code availability in their full text manuscripts. Of these, only 53 (70.7%) also reported this in their data availability statements (Table 4).

Table 4: Comparison of code availability declarations between data availability statements and full text manuscripts.

|  | | **Full text** | |
| --- | --- | --- | --- |
|  | **\_** | **Code mentioned** | **No code mentioned** |
| **Data availability statement** | **Code mentioned** | 53 | 16 |
| **No code mentioned** | 22 | 309 |

##### 

# Discussion

## Principal findings

We have reviewed 4101 preprinted and 159 published data availability statements, coding them as “open” or “closed” according to a predefined classification system. During this labor-intensive process, we appreciated statements that reflected the authors’ enthusiasm for data sharing (“YES”),[22] their bluntness (“Data is not available on request.”),[23] and their efforts to endear themselves to the reader (“I promise all data referred to in the manuscript are available.”).[24]

A minority of published records we examined contained a data availability statement (n = 159 (42.0%)). This lack of availability statement at publication results in a loss of useful information. For at least one published article, we identified relevant information in the preprinted statement that did not appear anywhere in the published article25,26]

However, simply being required to complete a data availability statement does not encourage authors to not make their data publicly available. This is illustrated by the fact that, at both the preprint and published stages, the most common category was “Data available on request” (Figure 1). In light of the substantial impact that studies published as preprints on medRxiv have had on real-time decision making during the current COVID-19 pandemic,[11] it is concerning that data for these preprints is so infrequently readily available for inspection.

We provide initial descriptive evidence that strict data-sharing policies, which require data to be made openly available (where appropriate) as a condition of publication, appear to succeed in making research data more open than those that simply encourage data sharing. Our findings, though based on a relatively small number of observations, agree with other studies on the effect of journal policies on author behavior. A recent randomized controlled trial found that encouraging authors to follow a reporting checklist (the ARRIVE checklist, designed to improve reporting) was not effective in increasing reporting quality, suggesting a strategy of encouragement is not enough to alter author behaviour.[27,28] Conversely, in cases where authors were required to add a data availability statement, studies have found significant increase in the number of authors completing the field.[2]

Finally, we also provide evidence that data availability statements alone are insufficient to capture code availability declarations in a standardized input field. Code sharing has been advocated strongly elsewhere,[29,30] as it provides an insight into the analytic decisions made by the research team, and there are few (if any) circumstances in which it is not possible to share the analytic code underpinning an analysis. Similar to data availability statements, a dedicated code availability statement, which is critically assessed as part of the publication process, will help researchers to assess published results.

## Strengths and limitations

A particular strength of this analysis is that the design allows us to compare what is essentially the same paper (same design, findings and authorship team) under two different data-sharing polices, and assess the change in the statement between them. To our knowledge this is the first study to use this approach to examine the impact of journal editorial policies, though a similar design was used to assess the change in reporting standards between preprinted and published bioRxiv articles.[31]

However, our analysis is subject to a number of potential limitations. The primary limitation is the fact that manuscripts (at both the preprint and published stages) may have included links to the data, or more information that uniquely identifies the dataset from a data portal, within the text (for example, in the Methods section). While this might be the case, if readers are expected to piece together the relevant information from different locations in the manuscript, it throws into question what having a dedicated data availability statement adds. Secondly, Category 1 (“Data not available”) is likely to include studies which justify the authors’ decision not to make data available, for example, if studying a rare disease, releasing the datasets would make the patients readily identifiable. We wish to stress that this is situation is perfectly reasonable, and that our aim is not to have all research data available openly. Rather, we aim to move towards informative and useful data availability statements, and to require authors to state explicitly why they have not shared their data. A final limitation is that we do not assess the veracity of any data availability statements, which may introduce some misclassification bias into our categorization. For example, we do not check whether all relevant data can actually be found in the manuscript/supplementary files (Category 7) or the linked repository (Category 8). Previous work has suggested that this is unlikely to be the case.[8]

## Implications for policy

Based on our analysis, journals that require data sharing statements publish a higher number of articles whose data is openly available. Moreover, where data is not available, the reason for this is clearly stated. This would suggest that data sharing could be immediately improved by journals becoming more stringent in their data availability policies.

Furthermore, journals requiring and regulating data availability statements will increase the importance for researchers of curating data and holding it to a standard capable of being shared. It would also highlight the urgency to define clear and robust regulations around data reuse.

## Conclusion

Data availability statements are an important tool in the fight to make studies more reproducible. We would strongly encourage all journals to take a moment this holiday season and examine their data availability guidelines – although it’s been said many times, many ways; open data helps me and you OR “heal the world, let us see your datasets next time”

##### 

# Back Matter

## Material available statement

All materials (data, code and supporting information) are available on request to one S. Claus, North Pole, (or alternatively can be found at <https://github.com/mcguinlu/data-availability-impact>, archived at time of submission on Zenodo (DOI:)).

## Contributor statement

All authors reviewed this manuscript before approving the final version. LAM is guarantor of the article.

**CReditT Taxonomy** Conceptualization: Luke A. McGuinness. Data Curation: Luke A. McGuinness. Formal Analysis: Luke A. McGuinness and Athena L. Sheppard.  
Investigation: Luke A. McGuinness and Athena L. Sheppard.  
Methodology: Luke A. McGuinness and Athena L. Sheppard.  
Project Administration: Luke A. McGuinness.  
Software: Luke A. McGuinness.  
Supervision: Luke A. McGuinness.  
Validation: Luke A. McGuinness and Athena L. Sheppard.  
Visualization: Luke A. McGuinness.  
Writing - Original Draft Preparation: Luke A. McGuinness.  
Writing - Review & Editing: Luke A. McGuinness and Athena L. Sheppard.

## Competing interests

We have read and understood BMJ policy on declaration of interests and declare LAM is supported by an National Institute for Health Research (NIHR) Doctoral Research Fellowship (DRF-2018-11-ST2-048). The views expressed in this article are those of the authors and do not necessarily represent those of the NHS, the NIHR, MRC, or the Department of Health and Social Care.

##### 

# Highlights

## What is already known on this topic

* Data sharing is increasingly seen as a core component of good research practice.
* Data availability statements are completed by researchers when required, but by themselves,do not encourage researchers to make their data publicly available.

## What this study adds

* Similar to published articles, the majority of preprints claim to make data “available on request”.
* Strict editorial policies that mandate data sharing (where appropriate) as a condition of publication appear to be effective in making research data available.
* Introducing a code availability statement will increase the findability of code.

##### 

# Bibliography

1 Krumholz HM. Why data sharing should be the expected norm. *BMJ (Clinical research ed)* 2015;**350**:h599. doi:[10.1136/bmj.h599](https://doi.org/10.1136/bmj.h599)

2 Federer LM, Belter CW, Joubert DJ *et al.* Data sharing in PLOS ONE: An analysis of Data Availability Statements. *PLOS ONE* 2018;**13**:e0194768. doi:[10.1371/journal.pone.0194768](https://doi.org/10.1371/journal.pone.0194768)

3 Krawczyk M, Reuben E. (Un)Available upon Request: Field Experiment on Researchers’ Willingness to Share Supplementary Materials. *Accountability in Research* 2012;**19**:175–86. doi:[10.1080/08989621.2012.678688](https://doi.org/10.1080/08989621.2012.678688)

4 Naudet F, Sakarovitch C, Janiaud P *et al.* Data sharing and reanalysis of randomized controlled trials in leading biomedical journals with a full data sharing policy: Survey of studies published in *The BMJ* and *PLOS Medicine*. *BMJ* 2018;k400. doi:[10.1136/bmj.k400](https://doi.org/10.1136/bmj.k400)

5 Vines TH, Albert AYK, Andrew RL *et al.* The Availability of Research Data Declines Rapidly with Article Age. *Current Biology* 2014;**24**:94–7. doi:[10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014)

6 Vasilevsky NA, Minnier J, Haendel MA *et al.* Reproducible and reusable research: Are journal data sharing policies meeting the mark? *PeerJ* 2017;**5**. doi:[10.7717/peerj.3208](https://doi.org/10.7717/peerj.3208)

7 Colavizza G, Hrynaszkiewicz I, Staden I *et al.* The citation advantage of linking publications to research data. *PLOS ONE* 2020;**15**:e0230416. doi:[10.1371/journal.pone.0230416](https://doi.org/10.1371/journal.pone.0230416)

8 Roche DG, Kruuk LEB, Lanfear R *et al.* Public Data Archiving in Ecology and Evolution: How Well Are We Doing? *PLOS Biology* 2015;**13**:e1002295. doi:[10.1371/journal.pbio.1002295](https://doi.org/10.1371/journal.pbio.1002295)

9 Tan SC, Flanagan D, Morris E *et al.* Research data repositories chosen by researchers across broad range of disciplines, from an analysis of 145,000 data availability statements. *Authorea* Published Online First: July 2020. doi:[10.22541/au.159422974.49069472](https://doi.org/10.22541/au.159422974.49069472)

10 Rawlinson C, Bloom T. New preprint server for medical research. *BMJ* 2019;**365**. doi:[10.1136/bmj.l2301](https://doi.org/10.1136/bmj.l2301)

11 Fraser N, Brierley L, Dey G *et al.* Preprinting a pandemic: The role of preprints in the COVID-19 pandemic. *bioRxiv* 2020;2020.05.22.111294. doi:[10.1101/2020.05.22.111294](https://doi.org/10.1101/2020.05.22.111294)

12 McGuinness LA, Sheppard AL. Protocol for a descriptive analysis of the data availability statements accompanying medRxiv preprints. 2020.

13 McGuinness LA, Schmidt L. Medrxivr: Accessing medRxiv data in r. 2020.<https://github.com/mcguinlu/medrxivr>

14 Wickham H. *Rvest: Easily harvest (scrape) web pages*. 2019. <https://CRAN.R-project.org/package=rvest>

15 Chamberlain S, Zhu H, Jahn N *et al.* *Rcrossref: Client for various ’crossref’ ’apis’*. 2020. <https://CRAN.R-project.org/package=rcrossref>

16 R Core Team. *R: A language and environment for statistical computing*. Vienna, Austria:: R Foundation for Statistical Computing 2019. <https://www.R-project.org/>

17 Gamer M, Lemon J, <puspendra.pusp22@gmail.com> IFPS. *Irr: Various coefficients of interrater reliability and agreement*. 2019. <https://CRAN.R-project.org/package=irr>

18 Wickham H, Hester J, Chang W. *Devtools: Tools to make developing r packages easier*. 2019. <https://CRAN.R-project.org/package=devtools>

19 Wickham H, François R, Henry L *et al.* *Dplyr: A grammar of data manipulation*. 2019. <https://CRAN.R-project.org/package=dplyr>

20 Chan C-h, Chan GC, Leeper TJ *et al.* *Rio: A swiss-army knife for data file i/o*. 2018.

21 Wickham H. *Stringr: Simple, consistent wrappers for common string operations*. 2019. <https://CRAN.R-project.org/package=stringr>

22 chen ling, Du X, Liu Y *et al.* Comparison of the Clinical Implications among Two Different Nutritional Indices in Hospitalized Patients with COVID-19. *medRxiv* Published Online First: May 2020. doi:[10.1101/2020.04.28.20082644](https://doi.org/10.1101/2020.04.28.20082644)

23 Hashmi M, Taqi A, Memon MI *et al.* A national landscaping survey of critical care services in hospitals accredited for training in a lower-middle income country: Pakistan. *medRxiv* Published Online First: April 2020. doi:[10.1101/2020.04.22.20071555](https://doi.org/10.1101/2020.04.22.20071555)

24 Peng L, Liu J, Xu W *et al.* 2019 Novel Coronavirus can be detected in urine, blood, anal swabs and oropharyngeal swabs samples. *medRxiv* Published Online First: February 2020. doi:[10.1101/2020.02.21.20026179](https://doi.org/10.1101/2020.02.21.20026179)

25 Martin J, Hosking G, Wadon M *et al.* A brief report: De novo copy number variants in children with attention deficit hyperactivity disorder. *medRxiv* Published Online First: December 2019. doi:[10.1101/2019.12.12.19014555](https://doi.org/10.1101/2019.12.12.19014555)

26 Martin J, Hosking G, Wadon M *et al.* A brief report: De novo copy number variants in children with attention deficit hyperactivity disorder. *Translational Psychiatry* 2020;**10**:135. doi:[10.1038/s41398-020-0821-y](https://doi.org/10.1038/s41398-020-0821-y)

27 Hair K, Macleod MR, Sena ES *et al.* A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Research Integrity and Peer Review* 2019;**4**:12. doi:[10.1186/s41073-019-0069-3](https://doi.org/10.1186/s41073-019-0069-3)

28 Kilkenny C, Browne WJ, Cuthill IC *et al.* Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLOS Biology* 2010;**8**:e1000412. doi:[10.1371/journal.pbio.1000412](https://doi.org/10.1371/journal.pbio.1000412)

29 Goldacre B, Morton CE, DeVito NJ. Why researchers should share their analytic code. *BMJ* 2019;l6365. doi:[10.1136/bmj.l6365](https://doi.org/10.1136/bmj.l6365)

30 Eglen SJ, Marwick B, Halchenko YO *et al.* Towards standard practices for sharing computer code and programs in neuroscience. *Nature neuroscience* 2017;**20**:770–3. doi:[10.1038/nn.4550](https://doi.org/10.1038/nn.4550)

31 Carneiro CFD, Queiroz VGS, Moulin TC *et al.* Comparing quality of reporting between preprints and peer-reviewed articles in the biomedical literature. *medRxiv* Published Online First: March 2019. doi:[10.1101/581892](https://doi.org/10.1101/581892)