# Front Matter

**Title:**   
All I want for Christmas is you(r data): descriptive analysis of the availability statements accompanying medRxiv preprints

**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)

**Funding**

LAM is supported by an 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.

**Keywords**

Preprints; Observational study; Reproducibility; Data sharing

##### 

# Abstract (250-300 words)

**Objective** To assess the distribution of data availability statements across varying categories of “openness”, and to assess whether strict 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 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 became more open between preprint and publication more frequently when the journal mandated open data sharing (open at preprint: 33.3%, open at publication: 61.4%) compared to when the journal did not mandate open data sharing (open at preprint: 20.2%, open at publication: 22.3%).

**Conclusion**

Data availability statements alone are not enough to ensure transparency of reporting and reproducibility of published results. Strict data availability policies that require data sharing (where possible and appropriate) as a condition of publication appear to be effective in making research data more open.

##### 

# Introduction

## Background

Data availability statements are intended to provide readers with important information about whether and where the data described in an academic paper are available, so that others can examine, verify and build on the results of that study. However, these statements, particularly those that purport to make data “available on request”, are arguably more frequently used as a source of nihilistic humor among academics, who recount stories of frustrating email exchanges with uncooperative primary study authors at the faculty Christmas party.

While several journals have implemented editorial policies that require authors to complete a data availability statement in the past decade,1 few have policies that require data sharing (where appropriate) as a condition of publication. medRxiv, the preprint repository for manuscripts in the medical, clinical, and related health sciences,2 has implemented a similar approach, requiring manuscript authors to complete a data availability statement in order to have their preprint accepted to the server but not performing any critical review of these statements. While making data availability statements compulsory is a step in the right direction, it arguably does not achieve it’s intended effect of making research data more open. Preprints, in particular those 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. As such, assessing the “openness” of preprint data availability statements in their own right is worthwhile. Additionally, comparing the preprint and published data availability statements for the same paper allows for the potential impact of journal data sharing policies to be examined.

Our primary aim was to explore the distribution of data availability statements across the categories 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

A protocol for this analysis was registered in advance and followed at all stages of the study.3 Any deviations from the protocol are described.

## 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.4,5 Details on the journal in which preprints were subsequently published was extracted using the published DOI provided by medRxiv and rcrossref.6 Several other R packages were used for data cleaning and analysis.7–12

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,13 the categories used by previous effort to categorize published data availability statements,1,14 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 Files XXXX 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.

To assess whether data availability statements change between preprint and published articles, we assessed whether a discrepancy existed between the categories assigned to the preprinted and published statements, and the direction of the discrepancy (i.e. does it become more closed or more open). We declare a minor deviation from our protocol,3 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”).

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).”f |
| **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] | | | |

##### 

# Results

The data availability statements accompanying 4101 preprints, registered on medRxiv between 25th June 2019 and 1st May 2020, were extracted from the medRxiv preprint repository on the 26th May 2020 and were categorized according to the categories in Table 1. Inter-rater reliability 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. 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.

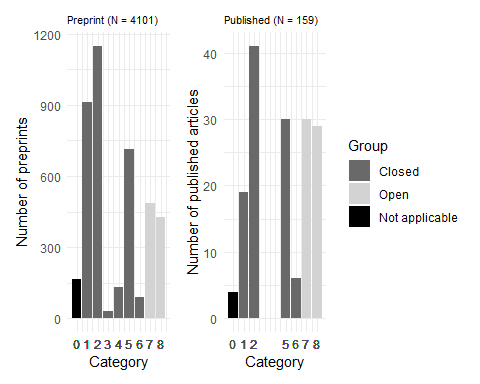


Figure 1: Distribution of the data availability statements of preprinted and published records by category. The categories from Table 1 are presented on the X-axis.

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).

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 data sharing** | 70 | 94 | 20.2% (19) | 22.3% (21) | 10 | 8 | 76 |
| **Requires data sharing** | 20 | 57 | 33.3% (19) | 61.4% (35) | 16 | 0 | 41 |

161 (3.9%) preprints stated that data would be available on publication (28 (0.7%) included a link to the embargoed repository, while 133 (3.2%) did not). However only 10 of these preprints have 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 (33.3%) |
| **Available in future (no link)** | 7 | 5 (71.4%) |

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

We have reviewed 4101 preprint data availability statements, coding them into varying categories of openness, and have sought to compare them to their published counterparts, where available. During this process, we appreciated statements that reflected the authors enthusiasm for the data sharing (“YES”15), their bluntness (“Data is not available on request.”16), and their efforts to endear themselves to the reader (“I promise all data referred to in the manuscript are available.”17).

A minority of published records we examined contained a data availability statement (159 (42.0%)). By comparing the published paper with its preprint, we can demonstrate that this results in important information being lost. As an example, we identified a preprint with an informative availability statement, including a link to the specific open-access dataset used,18 which has no corresponding data availability statement when published (the information about, and link to, the data is not found anywhere else in the published paper).19 A recent analysis have shown that when mandated, data availability statements are completed by authors,1 and so we recommend that completing this metadata should be compulsory across all journals.

While requiring authors to submit a data availability statement is a good first step, simply providing a text box for authors to complete without critical review will not move authors toward making their data 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). While this may seem like an acceptable manner in which to make research data available, there are a number of issues with this approach. Previous work has shown that these statements are demonstrably false in the majority of cases: that when data is requested, it is not actually made available.20 As an illustrative example, a recent editorial in Molecular Brain demonstrated that, of 41 manuscripts for which data was requested during editorial review, more than half withdrew their manuscript rather than provide the raw data.21 If this is the response from authors when there is a strong incentive to share their data (i.e. publication in the journal), we do not hold out hope that the response to requests from other researchers, which sometimes intend to critique the original study’s findings, will be better. Additionally, making data available on request opens up the possibility of differential responses based on requester characteristics (e.g. gender, career stage, or institution). Finally, previous work found that the availability of data “available on request” declines with article age, meaning that this approach is not a valid long term option for data sharing.22

We provide evidence that critical review of data availability statements, in line with data sharing policies that require data sharing (where appropriate) as a condition of publication, succeeds in making research data more open. As an example, of the four articles published in F1000Research, which has one of the strictest data-sharing policies (“Failure to openly provide data for publication without good justification is likely to result in your article being rejected.”), two records for which data were available on request at the preprint stage made their data openly available from an online repository on publication.23–26 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 simply requesting that authors follow a reporting checklist (the ARRIVE checklist) was not effective in increasing reporting quality.27,28 Similarly, a review of data availability statements published in high impact journals found that data availability statements frequently do not conform to the policies of the journals they are published in, indicating a lack of critical review of these statements.29

We also found that the datasets used in a study may be frequently be open access, but there is insufficient detail captured in the data availability statement, such as date of data extraction or a unique project/accession number, to identify the specific dataset used (Category 5, Figure 1). A primary example of this was the Johns Hopkins COVID-19dataset, which is updated daily.30 Several preprinted studies made use of this data in their analysis, and reported it as a data source in their data availability statements, but did not provide the date of access. Relatedly, the medRxiv preprint repository has grown explosively as a result of the COVID-19 pandemic. Close to half the titles and abstract of preprints in our sample (N = 2017) contained a term related to the COVID-19 pandemic (“NCOV”,"coronavir\*“,”COVID"). In light of this, and the substantial impact that studies published as preprints have had on real-time decision making during the current pandemic, it is concerning that data for medRxiv preprints is so infrequently available for inspection.

Finally, we also provide evidence that data availability statements alone are insufficient to capture the availability of code. Code sharing has been advocated strongly elsewhere,31,32 as 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, having a dedicated code availability statement which is critically assessed as part of the publication process will help to increase transparency of reporting and reproducibility of published results.

## Limitations

Our analysis is subject to a number of potential limitations, the primary of which is that manuscripts may have included links to the data, or more information that uniquely identifies the dataset from a data portal, within the text (e.g. in the Methods section). While this might be the case, it raises serious concerns about the purpose and usefulness of data availability statements. 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 to the manuscript.

A related limitation is that we chose to focus on the text of the data availability statement alone when making classifications, rather than incorporating context. A key example of this is when an “open” data availability statement is included in a closed access publication, or when the manuscript??. However, we feel this is a valid approach, and would advocate for data availability statements to be considered self-contained article meta-data, so that they can be access without having to pay for access, similar to the approach already taken by some Nature journals (as an example from our study data, see Yan 2020 33).

It is important to note that Category 1 (“Data not available”) is likely to include studies which have a good reason for not making their data available. As an 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 (as this will likely never be the case), but more to move towards informative and useful data availability statements, and to require authors to state explicitly why they have not shared their data. Finally, we do not assess the veracity of any data availability statements, meaning we do not verify that data can be found in the open access repository listed (Category 8), that a data availability statement is not applicable (Category 0) or that all data is included in the manuscript and supplementary materials (Category 7).

## Conclusion

Strict data availability policies, those that require data sharing (where possible and appropriate) as a condition of publication, appear to be effective in making research data more open. We recommend implementing critical editorial/peer review of the data availability statements submitted for publication with an aim to require data sharing, where possible, and where not, to produce informative statements that justify the lack of open data. In addition, we suggest that code availability statements (or a composite “Material” availability section) be introduced, as data availability statements alone do not adequately capture code availability declarations.

##### 

# Back Matter

## Material available statement

All materials (data, code and supporting information) are available on request to the corresponding author, following payment of one red-nosed reindeer (or alternatively can be found at <https://github.com/mcguinlu/data-availability-impact>).

##### 

# Highlights

## What is already known on this topic

Existing studies on DAS in published articles.

## What this study adds

No previous study has examined the distribution of data availability statements between preprint and publication?, and examined the association of “open” data availability statements with subsequent publication.

No previous study has examined the change in openness associated with the peer-review process, whether code

##### 

# Supplementary materials

## Decision rules for exceptional data availability statement

* For items that met the criteria for multiple categories (e.g. “Most of the data analyzed in this manuscript are provided either within the manuscript itself, or in the manuscript posted by Sasani et al. on bioRxiv at <https://www.biorxiv.org/content/10.1101/552117v2> and its accompanying links; additional data may be accessed by contacting the corresponding author (Dr. Cawthon).”), we took a conservative approach, and categorized them based on the least permissive aspect of the statement. This approach was taken on the basis that ***all*** data must be available - for example, in the above statement, the additional information that is available on request may be required in order to reproduce the analysis.

## 

**F1000Research**

**PLOS ONE**

##### 

# Bibliography

1. Federer LM, Belter CW, Joubert DJ, et al. Data sharing in PLOS ONE: An analysis of Data Availability Statements. Wicherts JM, ed. *PLOS ONE*. 2018;13(5):e0194768. doi:[10.1371/journal.pone.0194768](https://doi.org/10.1371/journal.pone.0194768).

2. 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).

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

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

5. Wickham H. *Rvest: Easily Harvest (Scrape) Web Pages*.; 2019. <https://CRAN.R-project.org/package=rvest>.

6. Chamberlain S, Zhu H, Jahn N, Boettiger C, Ram K. *Rcrossref: Client for Various ’Crossref’ ’Apis’*.; 2020. <https://CRAN.R-project.org/package=rcrossref>.

7. R Core Team. *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing; 2019. <https://www.R-project.org/>.

8. 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>.

9. Wickham H, Hester J, Chang W. *Devtools: Tools to Make Developing R Packages Easier*.; 2019. <https://CRAN.R-project.org/package=devtools>.

10. Wickham H, François R, Henry L, Müller K. *Dplyr: A Grammar of Data Manipulation*.; 2019. <https://CRAN.R-project.org/package=dplyr>.

11. Chan C-h, Chan GC, Leeper TJ, Becker J. *Rio: A Swiss-Army Knife for Data File I/O*.; 2018.

12. Wickham H. *Stringr: Simple, Consistent Wrappers for Common String Operations*.; 2019. <https://CRAN.R-project.org/package=stringr>.

13. Wilkinson MD, Dumontier M, Aalbersberg IJ, et al. The FAIR Guiding Principles for scientific data management and stewardship. *Scientific Data*. 2016;3(1):160018. doi:[10.1038/sdata.2016.18](https://doi.org/10.1038/sdata.2016.18).

14. Colavizza G, Hrynaszkiewicz I, Staden I, Whitaker K, McGillivray B. The citation advantage of linking publications to research data. *PLOS ONE*. 2020;15(4):e0230416. doi:[10.1371/journal.pone.0230416](https://doi.org/10.1371/journal.pone.0230416).

15. 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*. May 2020. doi:[10.1101/2020.04.28.20082644](https://doi.org/10.1101/2020.04.28.20082644).

16. 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*. April 2020. doi:[10.1101/2020.04.22.20071555](https://doi.org/10.1101/2020.04.22.20071555).

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

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

19. 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(1):135. doi:[10.1038/s41398-020-0821-y](https://doi.org/10.1038/s41398-020-0821-y).

20. Krawczyk M, Reuben E. (Un)Available upon Request: Field Experiment on Researchers’ Willingness to Share Supplementary Materials. *Accountability in Research*. 2012;19(3):175-186. doi:[10.1080/08989621.2012.678688](https://doi.org/10.1080/08989621.2012.678688).

21. Miyakawa T. No raw data, no science: Another possible source of the reproducibility crisis. *Molecular Brain*. 2020;13(1):24. doi:[10.1186/s13041-020-0552-2](https://doi.org/10.1186/s13041-020-0552-2).

22. Vines TH, Albert AYK, Andrew RL, et al. The Availability of Research Data Declines Rapidly with Article Age. *Current Biology*. 2014;24(1):94-97. doi:[10.1016/j.cub.2013.11.014](https://doi.org/10.1016/j.cub.2013.11.014).

23. Hsiang W, Forman H, Jain S, et al. COVID-19 testing capabilities at urgent care centers in states with greatest disease burden. *F1000Research*. 2020;9:328. doi:[10.12688/f1000research.23203.1](https://doi.org/10.12688/f1000research.23203.1).

24. Pais RJ, Taveira N. Predicting the evolution and control of the COVID-19 pandemic in Portugal. *F1000Research*. 2020;9:283. doi:[10.12688/f1000research.23401.1](https://doi.org/10.12688/f1000research.23401.1).

25. Hsiang WR, Forman HP, Jain S, et al. Scarce COVID-19 Testing Capabilities at Urgent Care Centers in States with Greatest Disease Burden. *medRxiv*. March 2020. doi:[10.1101/2020.03.22.20040923](https://doi.org/10.1101/2020.03.22.20040923).

26. Pais RJ, Taveira N. Predicting the evolution and control of COVID-19 pandemic in Portugal. *medRxiv*. March 2020. doi:[10.1101/2020.03.28.20046250](https://doi.org/10.1101/2020.03.28.20046250).

27. Hair K, Macleod MR, Sena ES, on behalf of the IICARus Collaboration. A randomised controlled trial of an Intervention to Improve Compliance with the ARRIVE guidelines (IICARus). *Research Integrity and Peer Review*. 2019;4(1):12. doi:[10.1186/s41073-019-0069-3](https://doi.org/10.1186/s41073-019-0069-3).

28. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG. Improving Bioscience Research Reporting: The ARRIVE Guidelines for Reporting Animal Research. *PLOS Biology*. 2010;8(6):e1000412. doi:[10.1371/journal.pbio.1000412](https://doi.org/10.1371/journal.pbio.1000412).

29. Alsheikh-Ali AA, Qureshi W, Al-Mallah MH, Ioannidis JPA. Public Availability of Published Research Data in High-Impact Journals. Boutron I, ed. *PLoS ONE*. 2011;6(9):e24357. doi:[10.1371/journal.pone.0024357](https://doi.org/10.1371/journal.pone.0024357).

30. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in real time. *The Lancet Infectious Diseases*. 2020;20(5):533-534. doi:[10.1016/S1473-3099(20)30120-1](https://doi.org/10.1016/S1473-3099(20)30120-1).

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

32. Eglen SJ, Marwick B, Halchenko YO, et al. Towards standard practices for sharing computer code and programs in neuroscience. *Nature neuroscience*. 2017;20(6):770-773. doi:[10.1038/nn.4550](https://doi.org/10.1038/nn.4550).

33. Yan Q, Weeks DE, Xin H, et al. Deep-learning-based prediction of late age-related macular degeneration progression. *Nature Machine Intelligence*. 2020;2(2):141-150. doi:[10.1038/s42256-020-0154-9](https://doi.org/10.1038/s42256-020-0154-9).