



## Original Investigation | Health Policy

# Evaluation of Data Sharing After Implementation of the International Committee of Medical Journal Editors Data Sharing Statement Requirement

Valentin Danchev, DPhil; Yan Min, MD; John Borghi, PhD; Mike Baiocchi, PhD; John P. A. Ioannidis, MD, DSc

## Abstract

**IMPORTANCE** The benefits of responsible sharing of individual-participant data (IPD) from clinical studies are well recognized, but stakeholders often disagree on how to align those benefits with privacy risks, costs, and incentives for clinical trialists and sponsors. The International Committee of Medical Journal Editors (ICMJE) required a data sharing statement (DSS) from submissions reporting clinical trials effective July 1, 2018. The required DSSs provide a window into current data sharing rates, practices, and norms among trialists and sponsors.

**OBJECTIVE** To evaluate the implementation of the ICMJE DSS requirement in 3 leading medical journals: *JAMA*, *Lancet*, and *New England Journal of Medicine (NEJM)*.

**DESIGN, SETTING, AND PARTICIPANTS** This is a cross-sectional study of clinical trial reports published as articles in *JAMA*, *Lancet*, and *NEJM* between July 1, 2018, and April 4, 2020. Articles not eligible for DSS, including observational studies and letters or correspondence, were excluded. A MEDLINE/PubMed search identified 487 eligible clinical trials in *JAMA* (112 trials), *Lancet* (147 trials), and *NEJM* (228 trials). Two reviewers evaluated each of the 487 articles independently.

**EXPOSURE** Publication of clinical trial reports in an ICMJE medical journal requiring a DSS.

**MAIN OUTCOMES AND MEASURES** The primary outcomes of the study were declared data availability and actual data availability in repositories. Other captured outcomes were data type, access, and conditions and reasons for data availability or unavailability. Associations with funding sources were examined.

**RESULTS** A total of 334 of 487 articles (68.6%; 95% CI, 64%-73%) declared data sharing, with nonindustry NIH-funded trials exhibiting the highest rates of declared data sharing (89%; 95% CI, 80%-98%) and industry-funded trials the lowest (61%; 95% CI, 54%-68%). However, only 2 IPD sets (0.6%; 95% CI, 0.0%-1.5%) were actually deidentified and publicly available as of April 10, 2020. The remaining were supposedly accessible via request to authors (143 of 334 articles [42.8%]), repository (89 of 334 articles [26.6%]), and company (78 of 334 articles [23.4%]). Among the 89 articles declaring that IPD would be stored in repositories, only 17 (19.1%) deposited data, mostly because of embargo and regulatory approval. Embargo was set in 47.3% of data-sharing articles (158 of 334), and in half of them the period exceeded 1 year or was unspecified.

**CONCLUSIONS AND RELEVANCE** Most trials published in *JAMA*, *Lancet*, and *NEJM* after the implementation of the ICMJE policy declared their intent to make clinical data available. However, a wide gap between declared and actual data sharing exists. To improve transparency and data reuse, journals should promote the use of unique pointers to data set location and standardized choices for embargo periods and access requirements.

JAMA Network Open. 2021;4(1):e2033972. doi:10.1001/jamanetworkopen.2020.33972

**Open Access.** This is an open access article distributed under the terms of the CC-BY License.

## Key Points

**Question** What are the rates of declared and actual sharing of clinical trial data after the medical journals' implementation of the International Committee of Medical Journal Editors data sharing statement requirement?

**Findings** In this cross-sectional study of 487 clinical trials published in *JAMA*, *Lancet*, and *New England Journal of Medicine*, 334 articles (68.6%) declared data sharing. Only 2 (0.6%) individual-participant data sets were actually deidentified and publicly available on a journal website, and among the 89 articles declaring that individual-participant data would be stored in secure repositories, data from only 17 articles were found in the respective repositories as of April 10, 2020.

**Meaning** These findings suggest that there is a wide gap between declared and actual sharing of clinical trial data.

+ [Invited Commentary](#)

+ [Supplemental content](#)

Author affiliations and article information are listed at the end of this article.

## Introduction

Responsible sharing of individual-participant data (IPD) from clinical studies has gained increasing traction and has been advocated for many years by many scientists and scientific leadership organizations.<sup>1</sup> However, promoting data sharing from clinical trials has not been straightforward, and there has been much debate surrounding privacy risks and the optimal incentives for clinical trialists and sponsors.<sup>2-6</sup> Recently, the International Committee of Medical Journal Editors (ICMJE) implemented a clinical trial data sharing policy. The policy does not mandate<sup>7,8</sup> data sharing but requires a data sharing statement (DSS) from submissions reporting clinical trials effective July 1, 2018.<sup>9-11</sup>

Prior work has identified a range of potential risks preventing trialists from sharing IPD.<sup>3,12,13</sup> These risks include protection of patient privacy and confidentiality,<sup>4,13</sup> inappropriate data reuse and replication,<sup>12</sup> and researchers' and sponsors' potential losses of secondary publications and product advantage, respectively, because of the use of the shared data by competitors.<sup>3,5,14</sup> Repositories for clinical data from industry-funded<sup>15-17</sup> and publicly funded<sup>18</sup> trials have provided a safeguarded mechanism for responsible IPD sharing, thereby substantially minimizing patient privacy and confidentiality risks. However, perceived risks of inappropriate reuse and competition have been difficult to mitigate, especially when the current reward system for researchers predominantly incentivizes high-impact publications, often based on exclusive data, at the expense of transparency, reproducibility, and data reuse.<sup>19-22</sup>

Disincentives for data sharing are known to have a disproportionate impact on clinical studies because the process of conducting those studies is time, cost, and labor intensive.<sup>3</sup> Yet the role of prevalent disincentives and incentives (eg, data authorship<sup>23,24</sup>) for clinical trial data sharing have only recently entered the public realm,<sup>3,23,25,26</sup> in part accelerated by discussions surrounding the ICMJE's data sharing policy<sup>11,27</sup> when many points of agreement and disagreement among stakeholders were articulated.<sup>3,5,6,27,28</sup>

Many data repositories have been established to facilitate secure sharing of IPD from clinical trials.<sup>15,17,18,29-32</sup> Some industry sponsors, such as GlaxoSmithKline and Johnson & Johnson, have initiated their own data sharing repositories and partnerships with ClinicalStudyDataRequest.com (CSDR)<sup>17</sup> and the Yale University Open Data Access (YODA) Project,<sup>15</sup> respectively. Foundations and research charities, such as Wellcome Trust, Cancer Research UK, the Bill and Melinda Gates Foundation, and the UK Medical Research Council, have also implemented clinical trial data sharing policies and are now members of the CSDR platform.<sup>5,33</sup> Information on CSDR, the YODA Project, and other clinical trial data repositories can be found in the eTable in the [Supplement](#).

Following the 2003 Data Sharing Policy<sup>34</sup> that encouraged data plan and sharing of grants exceeding \$500 000, the National Institutes of Health (NIH) have implemented disease-specific data repositories, including the Biologic Specimen and Data Repository Information Coordinating Center,<sup>18</sup> which is supported by the National Heart, Lung, and Blood Institute (NHLBI), and the NCTN/NCORP Data Archive for data sets from clinical trials of the National Clinical Trials Network (NCTN) and the NCI Community Oncology Research Program (NCORP), which is supported by the National Cancer Institute. Recently, the NIH announced new Policy for Data Management and Sharing<sup>35</sup> with direct implications for clinical trial data sharing.<sup>21</sup>

Prior research has identified differences in clinical trial data sharing across funders and sponsors.<sup>30,36,37</sup> Historically, sharing of clinical trial data has been more prevailing among industry funders and sponsors associated with data repositories. For example, for a subset of drugs and industry sponsors (ie, Roche, Lilly, Boehringer Ingelheim, and GlaxoSmithKline), Boutron et al<sup>30</sup> found that 53% (512 of 966 randomized clinical trials) of clinical trials registered on ClinicalTrials.gov were listed on CSDR, and for 40% (385 of 966 randomized clinical trials) of the trials all documents were available, including raw data sets. Academic research has been less engaged with clinical trial data sharing. For instance, a prominent NIH repository such as the NHLBI Data Repository contained only 100 studies as of May 31, 2016.<sup>18</sup> However, a study<sup>37</sup> of trialists' intentions to share IPD from

clinical trials registered on ClinicalTrials.gov between January 2016 and August 2017 found that NIH-funded trials were more likely to indicate data sharing intentions than industry-funded trials.

The ICMJE policy requires investigators to state whether they will share data (or not) while simultaneously providing an opportunity for them to place multiple restrictions and conditions regarding data access. Specifically, the DSS provides an opportunity for authors and sponsors to specify periods of data exclusivity or embargo. In addition, authors can specify in the DSS how the data will be made available, reasons for data availability or unavailability, and related preferences (for examples of DSSs, see eAppendix 1 in the [Supplement](#)). Thus, the DSSs, required by the ICMJE's policy, provide a window into data sharing norms, practices, and perceived risks among trialists and sponsors. We set out to evaluate how the ICMJE's data sharing policy has been implemented in 3 leading medical journals that are also member journals of ICMJE: *JAMA*,<sup>9</sup> *Lancet*,<sup>10</sup> and *New England Journal of Medicine (NEJM)*.<sup>11</sup>

## Methods

Because this study used publicly available data and did not involve human participants, institutional review board approval and informed consent were not sought, in accordance with 45 CFR §46. The Strengthening the Reporting of Observational Studies in Epidemiology (STROBE) reporting guideline<sup>38</sup> was used as a guideline for this observational study. In particular, the STROBE checklist for cross-sectional studies was used to ensure accurate reporting.

The inclusion criteria in this study were reports of clinical trials, published as articles in *JAMA*, *Lancet*, and *NEJM* between July 1, 2018, and April 4, 2020, and containing a DSS. Excluded from the study were publications that contained no DSS because the article was submitted prior to July 1, 2018; was an observational or other type of study that is different from clinical trial and is, therefore, not required to contain a DSS under the ICMJE policy; or was a letter, correspondence, or other type of publication that lacks a DSS.

A MEDLINE/PubMed search of clinical trials published in the 3 journals between July 1, 2018, and April 4, 2020, identified 629 potentially eligible articles; 486 of them included a DSS, whereas the others were either submitted before July 2018, were not a clinical trial, or were letters. One article published in 2020 met all other inclusion criteria but contained no DSS and was included in the study sample as not sharing data because articles published in 2020 were likely submitted after July 1, 2018, and were, therefore, required to contain a DSS. We conducted a cross-sectional observational study for all 487 articles in *JAMA* (112 articles), *Lancet* (147 articles), and *NEJM* (228 articles) (eFigure in the [Supplement](#)). Two reviewers evaluated each article independently. Discrepancies were resolved unanimously or by a third reviewer.

Data were classified as being available when authors answered yes (*JAMA* and *NEJM*) or gave an unstructured positive response (*Lancet*) to the data availability question. Information about data type, access, conditions, and reasons for data availability or unavailability were taken from the DSS (eAppendix 2 in the [Supplement](#)). We also compared declared with actual data availability in repositories by examining whether information about data and data themselves are available in the respective repository. For more details on the MEDLINE/PubMed search strategy, inclusion and exclusion criteria for the study, definition of variables, and data extraction, see eAppendix 3 in the [Supplement](#).

Drawing on prior research reporting differences in intention to share clinical trial data between industry-funded and nonindustry (including NIH)-funded clinical trials,<sup>37</sup> we classified funding sources as industry, nonindustry NIH, nonindustry non-NIH, and mixed. Industry refers to research funding from companies. Nonindustry NIH refers to research funding from the US NIH. Nonindustry non-NIH refers to research funding from foundations, trusts, associations, national institutes outside the US, and so forth. Mixed refers to any combination of the other research-funding categories.

## Statistical Analysis

We conducted a descriptive analysis of variables associated with data sharing by type of funding and publication journal. For the primary outcome variables, declared and actual data sharing, we report the 95% CIs determined by bootstrapping (100 000 iterations). The  $\chi^2$  test was used for comparing differences in prevalence of declared data sharing between types of funding. The 2-tailed Fisher exact test was used for comparing differences in prevalence of data availability in repositories between types of funding.  $P < .05$  was considered significant. To perform data analysis and to generate summary statistics and graphs, we used the Python programming language version 3.8.3 (Python Software Foundation), a Jupyter Notebook,<sup>39</sup> and the following libraries: SciPy,<sup>40</sup> Pandas,<sup>41</sup> NumPy,<sup>42</sup> Matplotlib,<sup>43</sup> Scikits-Bootstrap,<sup>44</sup> and Seaborn.<sup>45</sup> For the statistical tests, we used the Python package Statsmodels and R statistical software version 4.0.2 (R Project for Statistical Computing).

## Results

Overall, 334 of 487 articles (68.6%; 95% CI, 64%-73%) declared data sharing (**Table 1**). The prevalence of declared data sharing varied by funder type for all 487 articles ( $\chi^2_3 = 15.93$ ;  $P = .001$ ). Nonindustry NIH-funded trials had the highest rates of declared data sharing (89%; 95% CI, 80%-98%) and industry-funded trials had the lowest rates (61%; 95% CI, 54%-68%) (**Figure 1A**). The ranking of funders regarding declared data sharing is largely consistent across the 3 journals—NIH, nonindustry non-NIH, mixed, and industry—although the 95% CIs overlap (**Figure 1B**). No substantial changes in the prevalence of declared data sharing were observed over the span of the first 7 quarters of policy implementation (**Figure 1C**).

The presence of multiple articles from the same clinical trials would violate the assumption of independence and could also introduce social dependencies,<sup>46</sup> such as clustering by authors, funder, or institution. To address this issue, we identified 12 clusters of multiple publications for the same trial. Each cluster contained 2 or (in 1 instance) 3 articles that had the same declared data sharing and funding source and were associated with common authors and institutions. We treated each cluster as a single article observation (474 articles) and recomputed our results about declared data sharing by funding as a way of assessing the impact of clustering on our results. Results were qualitatively similar: 326 of 474 articles declared data sharing (68.8%; 95% CI, 64%-73%), and differences in declared data sharing by funding sources were similar to the ones in the entire data set: NIH (88.4%; 95% CI, 79%-98%), nonindustry non-NIH (73.5%; 95% CI, 67%-80%), mixed (65.4%; 95% CI, 55%-76%), and industry (61.2%; 95% CI, 54%-68%). Owing to the small effect of clustering of articles from the same trial, the subsequent analysis uses the entire data set of 487 articles.

Industry sponsors had the lowest rate of declared data sharing in our sample. To examine further, we sampled industry sponsors that are members of the largest industry-initiated registry, CSDR, as of November 8, 2020, and identified the following industry sponsors with 1 or more studies in our sample: GlaxoSmithKline (10 studies), Novartis (10 studies), Bayer (6 studies), ViiV Healthcare (3 studies), ONO Pharmaceutical (2 studies), Sanofi (2 studies), Eisai (2 studies), Astellas (1 study), and Chugai (1 study). All of those industry sponsors are listed as current members of CSDR; industry sponsors that declared in their DSS that they would deposit data in CSDR but were not listed as members of the registry were excluded from this analysis. The rate of declared data sharing for all industry members of CSDR (56.8%; 95% CI, 38%-70%) was similar to the rate we established for all industry-funded trials (61.3%; 95% CI, 54%-68%). We could, therefore, exclude the possibility that the lowest rate of declared data sharing of industry funders is due to differences between companies that are members of data repositories and companies that are not members of data repositories.

Regarding type of shared data, 76.3% (255 of 334) of the articles proposed to provide deidentified IPD, 22.8% (76 of 334) would provide unspecified or partial data, and 0.9% (3 of 334) would provide aggregate data only. Only 2 of 334 IPD sets (0.6%; 95% CI, 0.0%-1.5%) were actually deidentified and publicly available (on journal website) as of April 10, 2020 (both data sets were

associated with the same clinical trial). The remaining were supposedly accessible via request to authors (143 of 334 articles [42.8%]), repository (89 of 334 articles [26.6%]), and company (78 of 334 articles [23.4%]) overall; 72 of 114 articles [63.2%] among industry-funded trials).

Conditions for access to data included embargo (158 of 334 articles [47.3%]), product approval (37 of 334 articles [11.1%]), and collaboration (9 of 334 articles [2.7%]). Among the 158 articles specifying embargo, approximately one-half required 1 year or less of data exclusivity. In the other half of embargo cases, the embargo period exceeded 1 year or was unspecified.

Table 1. Prevalence and Conditions of Declared Clinical Trial Data Sharing by Type of Funding

Variable	Articles, No./Total (%)				
	Industry (n = 186)	Nonindustry (n = 45)	Nonindustry, non-NIH (n = 173)	Mixed (n = 83)	Total (N = 487)
Declared data sharing, articles, No. (%) <sup>a</sup>					
Yes	114 (61.3)	40 (88.9)	127 (73.4)	53 (63.9)	334 (68.6)
No	72 (38.7)	5 (11.1)	46 (26.6)	30 (36.1)	153 (31.4)
Journal, articles, No. (%)					
JAMA	21 (11.3)	18 (40.0)	50 (28.9)	23 (27.7)	112 (23.0)
Lancet	57 (30.6)	7 (15.6)	64 (37.0)	19 (22.9)	147 (30.2)
New England Journal of Medicine	108 (58.1)	20 (44.4)	59 (34.1)	41 (49.4)	228 (46.8)
Type of declared available data <sup>b</sup>					
Deidentified individual-participant data	83/114 (72.8)	33/40 (82.5)	95/127 (74.8)	44/53 (83.0)	255/334 (76.3)
Aggregate data only	2/114 (1.8)	0/40 (0.0)	1/127 (0.8)	0/53 (0.0)	3/334 (0.9)
Unspecified or partial data <sup>c</sup>	29/114 (25.4)	7/40 (17.5)	31/127 (24.4)	9/53 (17.0)	76/334 (22.8)
Access to data <sup>d</sup>					
Request to authors	9/114 (7.9)	15/40 (37.5)	96/127 (75.6)	23/53 (43.4)	143/334 (42.8)
Request to committee, group, or unit	15/114 (13.2)	3/40 (7.5)	36/127 (28.3)	13/53 (24.5)	67/334 (20.1)
Request to repository or archive	46/114 (40.4)	22/40 (55.0)	5/127 (3.9)	16/53 (30.2)	89/334 (26.6)
Request to company	72/114 (63.2)	0/40 (0.0)	0/127 (0.0)	6/53 (11.3)	78/334 (23.4)
Access unspecified	5/114 (4.4)	2/40 (5.0)	5/127 (3.9)	3/53 (5.7)	15/334 (4.5)
Data are available to others	0/114 (0.0)	0/40 (0.0)	2/127 (1.6)	0/53 (0.0)	2/334 (0.6)
Conditional data access <sup>d</sup>					
Data embargo	49/114 (43.0)	22/40 (55.0)	57/127 (44.9)	30/53 (56.6)	158/334 (47.3)
Up to 1 y <sup>e</sup>	20/114 (17.5)	13/40 (32.5)	30/127 (23.6)	18/53 (34.0)	81/334 (24.3)
>1 y to 2 y	13/114 (11.4)	7/40 (17.5)	13/127 (10.2)	6/53 (11.3)	39/334 (11.7)
>2 y	4/114 (3.5)	2/40 (5.0)	4/127 (3.1)	5/53 (9.4)	15/334 (4.5)
Product approval	36/114 (31.6)	0/40 (0.0)	0/127 (0.0)	1/53 (1.9)	37/334 (11.1)
Collaboration	0/114 (0.0)	2/40 (5.0)	6/127 (4.7)	1/53 (1.9)	9/334 (2.7)
Reasons for why data not available <sup>d</sup>					
No reason given	40/72 (55.6)	2/5 (40.0)	26/46 (56.5)	15/30 (50.0)	83/153 (54.2)
Data privacy	3/72 (4.2)	1/5 (20.0)	9/46 (19.6)	2/30 (6.7)	15/153 (9.8)
Time and cost	0/72 (0.0)	0/5 (0.0)	2/46 (4.3)	0/30 (0.0)	2/153 (1.3)
Ongoing trial or research	2/72 (2.8)	0/5 (0.0)	2/46 (4.3)	9/30 (30.0)	13/153 (8.5)
Regulatory approval	3/72 (4.2)	0/5 (0.0)	1/46 (2.2)	0/30 (0.0)	4/153 (2.6)
Proprietary data	7/72 (9.7)	0/5 (0.0)	4/46 (8.7)	3/30 (10.0)	14/153 (9.2)
Shared among coinvestigators only	2/72 (2.8)	0/5 (0.0)	0/46 (0.0)	0/30 (0.0)	2/153 (1.3)
Data may be available for collaboration	6/72 (8.3)	0/5 (0.0)	0/46 (0.0)	1/30 (3.3)	7/153 (4.6)
Data may be available upon request	19/72 (26.4)	2/5 (40.0)	6/46 (13.0)	7/30 (23.3)	34/153 (22.2)

<sup>a</sup> The denominators represent the total for the respective column, unless otherwise indicated.

<sup>b</sup> Detailed definitions of variables are provided in eAppendix 2 in the [Supplement](#).

<sup>c</sup> Among the articles not specifying the type of data, 13 intended to store data in clinical trial data repositories, presumably individual-participant data, and 1 made available (via collaboration) individual-participant data in a designated repository.

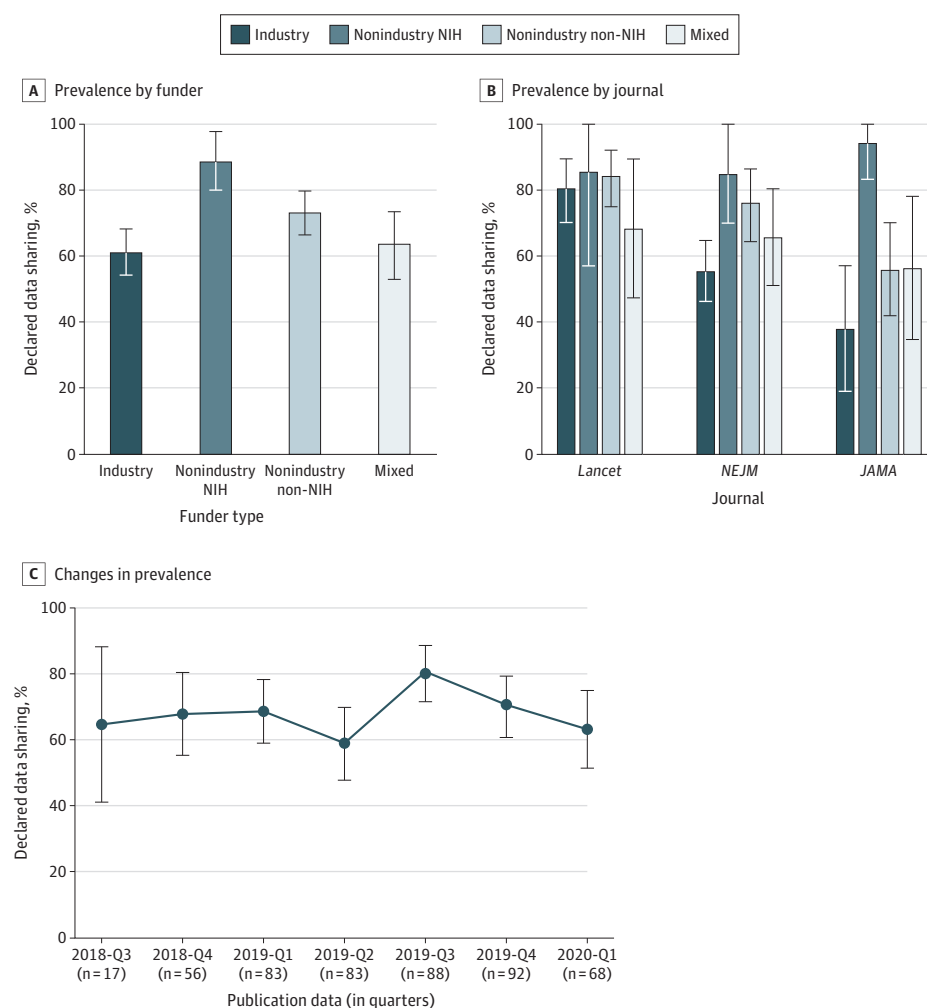
<sup>d</sup> Categories are not mutually exclusive.

<sup>e</sup> Time periods are not specified in all articles proposing data embargo.

Data repositories have a central role in improving sharing, security, discoverability, and reuse of research data,<sup>47,48</sup> particularly IPD from clinical trials.<sup>29,36,49</sup> Among the 89 articles proposing to make IPD available through repositories, many planned to store data in general-purpose repositories, including the CSDR (31 articles), the YODA Project (7 articles), and Vivli (7 articles). Another 30 articles planned to store IPD in NIH-supported, domain-specific data repositories, such as the NCTN/NCORP Data Archive (10 articles), the NHLBI Biologic Specimen and Data Repository Information Coordinating Center (9 articles), and the National Institute of Child Health and Human Development Data and Specimen Hub (5 articles) (**Figure 2**).

We compared declared with actual data availability in repositories (**Table 2**). Among 89 articles, information about the data was infrequently available in the repository (20 of 89 articles [22.5%]) and the data themselves were even more infrequently available there (17 of 89 articles [19.1%]) (**Figure 3**). Although data of NIH-funded trials (7 of 22 articles [31.8%]) were somewhat more likely to be available in repositories than data from industry-funded trials (7 of 46 articles [15.2%]), nonindustry non-NIH trials (0 of 5 articles), and trials with mixed funding (3 of 16 articles [18.8%]), data availability in repositories was not associated with type of funding ( $P = .32$ , Fisher exact test). Most trials provided neither information nor data in the respective repositories, mostly because of embargo and pending regulatory approval. Specifically, among the 72 articles that declared their intent but did not store data on repository, 37 (51%) made data access conditional on embargo or product approval.

Figure 1. Declared Clinical Trial Data Sharing in 3 Leading Medical Journals



Graphs show prevalence of data sharing by funder (A) and journal (B) and changes in prevalence between 2018 and 2020 (C). Error bars denote 95% CIs. *NEJM* indicates *New England Journal of Medicine*; NIH, National Institutes of Health.



Among reasons for data withholding, articles referred to data privacy (15 of 153 articles [9.8%]), proprietary data (14 of 153 articles [9.2%]), ongoing research (13 of 153 articles [8.5%]), and pending regulatory approval (4 of 153 articles [2.6%]). Most articles withholding data (83 of 153 articles [54.2%]) provided no reason. One article, which had no intent to make data available to others, was retracted.

Discussion

Most trials published in *JAMA*, *Lancet*, and *NEJM* after the endorsement of the ICMJE policy declared their intent to make clinical data available. Non-industry-funded trials communicated greater intent to share data than industry-funded trials, which exhibited low declared data sharing rates even among industry funders that are members of the largest data sharing repository, the CSDR. This result is consistent with prior research on intention to share data at the ClinicalTrials.gov<sup>37</sup> but departs from a trend among industry sponsors to establish mechanisms and repositories for sharing of clinical trial data.<sup>15,17,30,31,50</sup>

The commitment to data sharing substantially decreases when we consider indicators of actual vs declared data sharing: of 334 articles declaring that they would share data, only 2 IPD sets (0.6%; 95% CI, 0.0%-1.5%) were actually deidentified and publicly available on the journal website. Among the 89 articles declaring that they would store IPD in repositories, data from only 17 articles were found in the respective repository (Figure 3).

Although it is encouraging that data sharing appears widespread as a research norm<sup>51</sup> among trialists and that repositories for secure data sharing are often considered in DSSs, the low rate of actual data availability we identified is concerning. Some data sets that are currently unavailable may simply require additional time to be released, particularly those that are associated with embargo periods, but for many unavailable data sets, actionable information about availability is lacking. This points to the need for detailed requirements that encourage authors to engage with tangible and

Figure 2. Ranking of Data Repositories by the Number of Articles Intending to Share Individual-Participant Data (IPD) in the Respective Repository

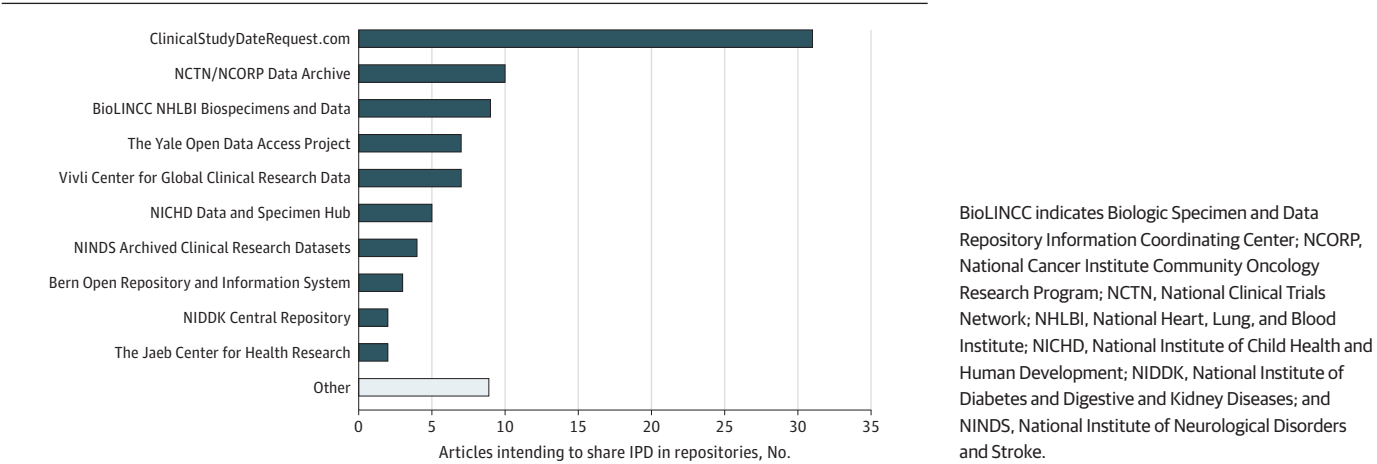


Table 2. Availability of Individual-Participant Data in Repository by Type of Funding

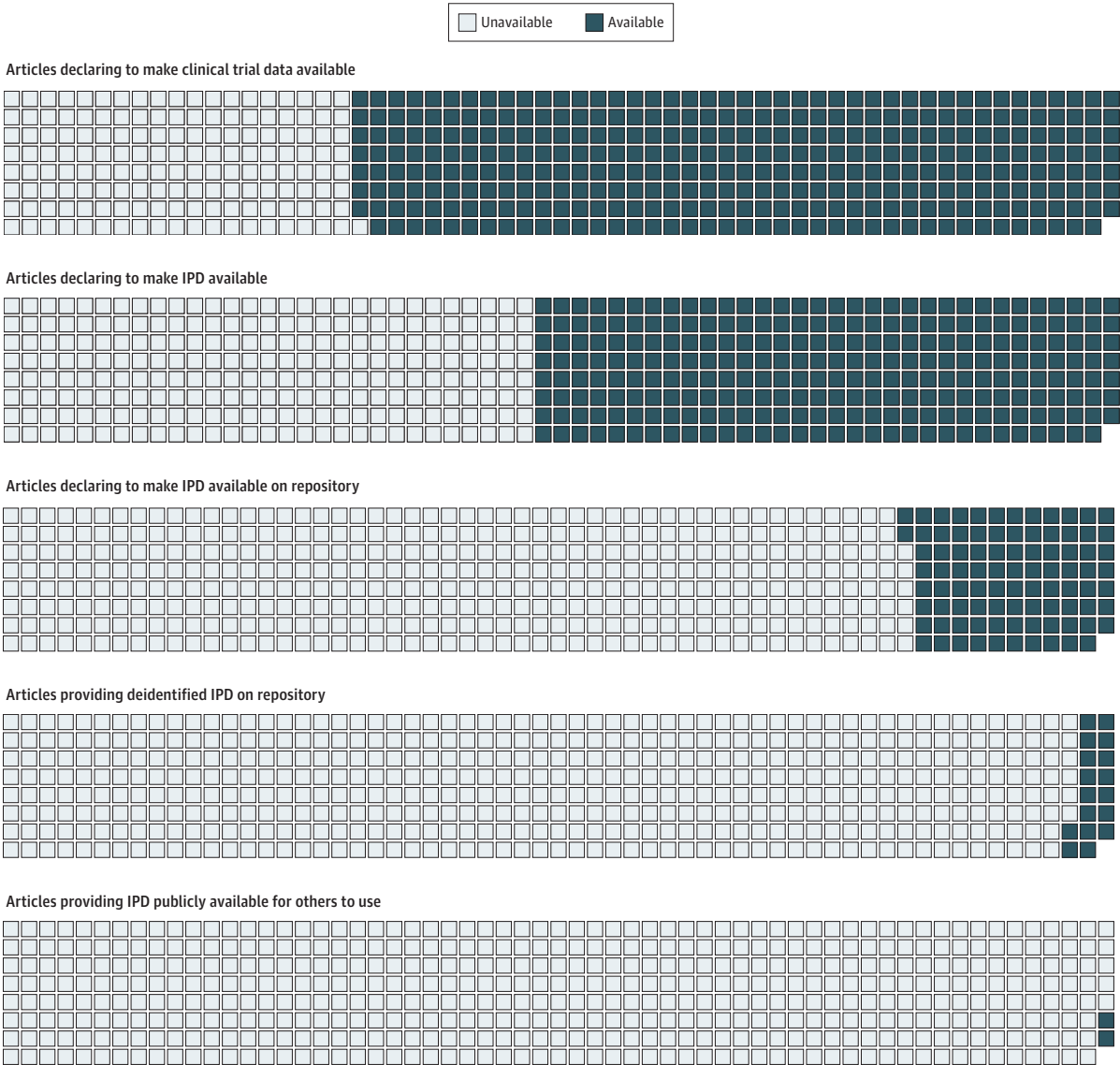
Variable	Articles, No. (%)				
	Industry (n = 46)	Nonindustry NIH (n = 22)	Nonindustry, non-NIH (n = 5)	Mixed (n = 16)	Total (N = 89)
Information about data is logged in repository	9 (19.6)	7 (31.8)	1 (20.0)	3 (18.8)	20 (22.5)
Data are available on repository to request	7 (15.2)	7 (31.8)	0 (0.0)	3 (18.8)	17 (19.1)

verifiable steps toward data sharing at or before publication, including logging information about the data sets in a data repository and having time-stamped information about their planned future release. Stakeholders, including journals, funders, and research institutions, should specifically focus on policies that could narrow the wide gap we identified between declared data sharing and actual availability of clinical trial IPD.

Consistent with prior research of clinical trial data registries<sup>30,37</sup> and DSSs,<sup>7</sup> the language of DSSs was often ambivalent. Offering of aggregate data, collaboration demands, lengthy or unspecified embargo periods, and the use of legacy methods for access such as author or company request communicate only lukewarm commitment. Repositories can be instrumental for sharing, but real practices may diverge from intent.

The current ICMJE requirement for DSSs is restricted to clinical studies. The policy restriction may be unwarranted given the high-profile observational studies recently retracted in *Lancet*<sup>52</sup> and *NEJM*<sup>53</sup> because of concerns regarding data veracity and availability, pointing to potential benefits of extending data sharing requirements to observational studies.<sup>54-56</sup>

Figure 3. Indicators of Declared and Actual Clinical Trial Individual-Participant Data (IPD) Availability as of April 10, 2020





## Limitations

Our study has limitations that should be acknowledged. First, only 3 journals were considered. Moreover, we could readily investigate declared vs actual data sharing practices only for repositories. Furthermore, only 2 IPD sets were deidentified and available on journal websites, so we could not meaningfully examine the usability of shared data or reproducibility<sup>8</sup> of the clinical trial studies. As more IPD sets become available, it would be interesting to assess whether they are easy to use, and how complete is the information being provided. More broadly, a repeated evaluation of data sharing intentions and practices could be valuable, particularly in the context of the ongoing clinical trial research response to the coronavirus disease 2019 pandemic that may have changed some norms and practices of clinical research.<sup>56-59</sup> In addition, we determined that repeated articles from the same clinical trial have very small effect on our estimates but did not examine clustering by authors and institutions that could occur beyond the level of single trials.

## Conclusions

To promote transparency and data reuse, journals and funders should work toward incentivizing data sharing via funding mechanisms<sup>21</sup> and data authorship,<sup>23</sup> and simultaneously discourage ambivalent wording in DSSs and possibly mandate data sharing. They can promote the use of unique pointers to data set location in repositories and to data request forms. Standardized choices for embargo periods, access requirements, and conditions for data use as part of the data sharing process could also reduce unnecessary data withholding and turn declarative data sharing into actual transparency in clinical trial data.

## ARTICLE INFORMATION

**Accepted for Publication:** November 29, 2020.

**Published:** January 28, 2021. doi:10.1001/jamanetworkopen.2020.33972

**Open Access:** This is an open access article distributed under the terms of the [CC-BY License](#). © 2021 Danchev V et al. *JAMA Network Open*.

**Corresponding Author:** Valentin Danchev, DPhil, Department of Sociology, University of Essex, Wivenhoe Park, Colchester CO4 3SQ, United Kingdom ([valentin.danchev@essex.ac.uk](mailto:valentin.danchev@essex.ac.uk)).

**Author Affiliations:** Meta-Research Innovation Center at Stanford, Stanford University School of Medicine, Stanford, California (Danchev, Ioannidis); Stanford Prevention Research Center, Department of Medicine, Stanford University School of Medicine, Stanford, California (Danchev, Ioannidis); Now with Department of Sociology, University of Essex, Colchester, United Kingdom (Danchev); Department of Epidemiology and Population Health, Stanford University School of Medicine, Stanford, California (Min, Baiocchi, Ioannidis); Lane Medical Library, Stanford University School of Medicine, Stanford, California (Borghi); Department of Biomedical Data Science, Stanford University School of Medicine, Stanford, California (Ioannidis).

**Author Contributions:** Dr Danchev had full access to all of the data in the study and takes responsibility for the integrity of the data and the accuracy of the data analysis.

*Concept and design:* Danchev, Min, Ioannidis.

*Acquisition, analysis, or interpretation of data:* Danchev, Min, Borghi, Baiocchi.

*Drafting of the manuscript:* Danchev, Borghi, Ioannidis.

*Critical revision of the manuscript for important intellectual content:* All authors.

*Statistical analysis:* Danchev, Baiocchi, Ioannidis.

*Obtained funding:* Ioannidis.

*Administrative, technical, or material support:* Min, Borghi.

*Supervision:* Ioannidis.

**Conflict of Interest Disclosures:** None reported.

**Funding/Support:** The Meta-Research Innovation Center at Stanford is supported by a grant from the Laura and John Arnold Foundation.

**Role of the Funder/Sponsor:** The funder had no role in the design and conduct of the study; collection, management, analysis, and interpretation of the data; preparation, review, or approval of the manuscript; and decision to submit the manuscript for publication.

**Additional Information:** All the data, computer code, and materials for this study are publicly available from the Open Science Framework at <https://osf.io/s5vbg/>.

## REFERENCES

1. Institute of Medicine. *Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk*. The National Academies Press; 2015.
2. Bauchner H, Golub RM, Fontanarosa PB. Data sharing: an ethical and scientific imperative. *JAMA*. 2016;315(12):1237-1239. doi:10.1001/jama.2016.2420
3. Devereaux PJ, Guyatt G, Gerstein H, Connolly S, Yusuf S; International Consortium of Investigators for Fairness in Trial Data Sharing. Toward fairness in data sharing. *N Engl J Med*. 2016;375(5):405-407. doi:10.1056/NEJMp1605654
4. Ursin G, Malila N, Chang-Claude J, et al. Sharing data safely while preserving privacy. *Lancet*. 2019;394(10212):1902. doi:10.1016/S0140-6736(19)32603-0
5. Kiley R, Peatfield T, Hansen J, Reddington F. Data sharing from clinical trials: a research funder's perspective. *N Engl J Med*. 2017;377(20):1990-1992. doi:10.1056/NEJMs1708278
6. Mello MM, Lieou V, Goodman SN. Clinical trial participants' views of the risks and benefits of data sharing. *N Engl J Med*. 2018;378(23):2202-2211. doi:10.1056/NEJMs1713258
7. Rowhani-Farid A, Barnett AG. Has open data arrived at the *British Medical Journal (BMJ)*? an observational study. *BMJ Open*. 2016;6(10):e011784. doi:10.1136/bmjopen-2016-011784
8. 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;360:k400. doi:10.1136/bmj.k400
9. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *JAMA*. 2017;317(24):2491-2492. doi:10.1001/jama.2017.6514
10. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *Lancet*. 2017;389(10086):e12-e14. doi:10.1016/S0140-6736(17)31282-5
11. Taichman DB, Sahni P, Pinborg A, et al. Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors. *N Engl J Med*. 2017;376(23):2277-2279. doi:10.1056/NEJMe1705439
12. Rathi V, Dzara K, Gross CP, et al. Sharing of clinical trial data among trialists: a cross sectional survey. *BMJ*. 2012;345:e7570. doi:10.1136/bmj.e7570
13. Berlin JA, Morris S, Rockhold F, Askie L, Ghersi D, Waldstreicher J. Bumps and bridges on the road to responsible sharing of clinical trial data. *Clin Trials*. 2014;11(1):7-12. doi:10.1177/1740774513514497
14. Van Noorden R. Data-sharing: everything on display. *Nature*. 2013;500(7461):243-245. doi:10.1038/nj7461-243a
15. Ross JS, Waldstreicher J, Bamford S, et al. Overview and experience of the YODA Project with clinical trial data sharing after 5 years. *Sci Data*. 2018;5(1):180268. doi:10.1038/sdata.2018.268
16. Strom BL, Buyse ME, Hughes J, Knoppers BM. Data sharing: is the juice worth the squeeze? *N Engl J Med*. 2016;375(17):1608-1609. doi:10.1056/NEJMp1610336
17. Rockhold F, Nisen P, Freeman A. Data sharing at a crossroads. *N Engl J Med*. 2016;375(12):1115-1117. doi:10.1056/NEJMp1608086
18. Coady SA, Mensah GA, Wagner EL, Goldfarb ME, Hitchcock DM, Giffen CA. Use of the National Heart, Lung, and Blood Institute Data Repository. *N Engl J Med*. 2017;376(19):1849-1858. doi:10.1056/NEJMs1603542
19. Moher D, Naudet F, Cristea IA, Miedema F, Ioannidis JPA, Goodman SN. Assessing scientists for hiring, promotion, and tenure. *PLoS Biol*. 2018;16(3):e2004089. doi:10.1371/journal.pbio.2004089
20. Nosek BA, Alter G, Banks GC, et al. Promoting an open research culture. *Science*. 2015;348(6242):1422-1425. doi:10.1126/science.aab2374

21. Sim I, Stebbins M, Bierer BE, et al. Time for NIH to lead on data sharing. *Science*. 2020;367(6484):1308-1309. doi:10.1126/science.aba4456
22. Alberts B, Cicerone RJ, Fienberg SE, et al. Self-correction in science at work. *Science*. 2015;348(6242):1420-1422. doi:10.1126/science.aab3847
23. Pierce HH, Dev A, Statham E, Bierer BE. Credit data generators for data reuse. *Nature*. 2019;570(7759):30-32. doi:10.1038/d41586-019-01715-4
24. Bierer BE, Crosas M, Pierce HH. Data authorship as an incentive to data sharing. *N Engl J Med*. 2017;376(17):1684-1687. doi:10.1056/NEJMs1616595
25. National Institutes of Health. Request for public comments on a draft NIH policy for data management and sharing and supplemental draft guidance (notice No. NOT-OD-20-013). Published November 6, 2019. Accessed December 9, 2020. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-20-013.html>
26. Tudur Smith C, Nevitt S, Appelbe D, et al. Resource implications of preparing individual participant data from a clinical trial to share with external researchers. *Trials*. 2017;18(1):319. doi:10.1186/s13063-017-2067-4
27. Taichman DB, Backus J, Baethge C, et al. Sharing clinical trial data: a proposal from the International Committee of Medical Journal Editors. *N Engl J Med*. 2016;374(4):384-386. doi:10.1056/NEJMe1515172
28. Rosenbaum L. Bridging the data-sharing divide: seeing the devil in the details, not the other camp. *N Engl J Med*. 2017;376(23):2201-2203. doi:10.1056/NEJMp1704482
29. Banzi R, Canham S, Kuchinke W, Krleza-Jeric K, Demotes-Mainard J, Ohmann C. Evaluation of repositories for sharing individual-participant data from clinical studies. *Trials*. 2019;20(1):169. doi:10.1186/s13063-019-3253-3
30. Boutron I, Dechartres A, Baron G, Li J, Ravaud P. Sharing of data from industry-funded registered clinical trials. *JAMA*. 2016;315(24):2729-2730. doi:10.1001/jama.2016.6310
31. Nisen P, Rockhold F. Access to patient-level data from GlaxoSmithKline clinical trials. *N Engl J Med*. 2013;369(5):475-478. doi:10.1056/NEJMs1302541
32. Li R, Scott J, Rockhold F, et al. Moving data sharing forward: the launch of the Vivli platform. Published November 19, 2018. Accessed December 9, 2020. <https://nam.edu/moving-data-sharing-forward-the-launch-of-the-vivli-platform/>
33. Humphreys G, Merriott G, Knowles R, Pierson B, Quattroni P. Clinical trial data sharing: what we've heard from researchers. Published February 20, 2020. Accessed December 9, 2020. [https://wellcome.figshare.com/articles/online\\_resource/Clinical\\_Trial\\_Data\\_Sharing\\_What\\_We\\_ve\\_Heard\\_from\\_Researchers/11847336](https://wellcome.figshare.com/articles/online_resource/Clinical_Trial_Data_Sharing_What_We_ve_Heard_from_Researchers/11847336)
34. National Institutes of Health. Final NIH statement on sharing research data. Published February 26, 2003. Accessed December 9, 2020. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-03-032.html>
35. National Institutes of Health. Final NIH policy for data management and sharing (notice No. NOT-OD-21-013). Published October 29, 2020. Accessed December 9, 2020. <https://grants.nih.gov/grants/guide/notice-files/NOT-OD-21-013.html>
36. Navar AM, Pencina MJ, Rymer JA, Louzao DM, Peterson ED. Use of open access platforms for clinical trial data. *JAMA*. 2016;315(12):1283-1284. doi:10.1001/jama.2016.2374
37. Bergeris A, Tse T, Zarin DA. Trialists' intent to share individual participant data as disclosed at ClinicalTrials.gov. *JAMA*. 2018;319(4):406-408. doi:10.1001/jama.2017.20581
38. von Elm E, Altman DG, Egger M, Pocock SJ, Gøtzsche PC, Vandenbroucke JP; STROBE Initiative. The Strengthening of Reporting of Observational Studies in Epidemiology (STROBE) statement: guidelines for reporting observational studies. *Ann Intern Med*. 2007;147(8):573-577. doi:10.7326/0003-4819-147-8-200710160-00010
39. Kluyver T, Ragan-Kelley B, Pérez F, et al. Jupyter Notebooks: a publishing format for reproducible computational workflows. In: Loizides F, Schmidt B, eds. *Positioning and Power in Academic Publishing: Players, Agents and Agendas*. IOS Press; 2016:87-90. doi:10.3233/978-1-61499-649-1-87
40. Virtanen P, Gommers R, Oliphant TE, et al; SciPy 1.0 Contributors. SciPy 1.0: fundamental algorithms for scientific computing in Python. *Nat Methods*. 2020;17(3):261-272. doi:10.1038/s41592-019-0686-2
41. McKinney W. Data structures for statistical computing in Python. Published 2010. Accessed December 9, 2020. <http://conference.scipy.org/proceedings/scipy2010/pdfs/mckinney.pdf>
42. van der Walt S, Colbert SC, Varoquaux G. The NumPy array: a structure for efficient numerical computation. *Comput Sci Eng*. 2011;13(2):22-30. doi:10.1109/MCSE.2011.37
43. Hunter JD. Matplotlib: a 2D graphics environment. *Comput Sci Eng*. 2007;9(3):90-95. doi:10.1109/MCSE.2007.55

44. Python Software. Scikits-Bootstrap version 1.0.1. Published November 20, 2020. Accessed December 9, 2020. <https://pypi.org/project/scikits.bootstrap/>
45. Zenodo. mwaskom/seaborn version 0.10.1. Published April 26, 2020. Accessed December 9, 2020. <https://zenodo.org/record/3767070#.X9F559hKg2w>
46. Danchev V, Rzhetsky A, Evans JA. Centralized scientific communities are less likely to generate replicable results. *Elife*. 2019;8:e43094. doi:10.7554/eLife.43094
47. Federer LM, Belter CW, Joubert DJ, et al. Data sharing in PLOS ONE: an analysis of data availability statements. *PLoS One*. 2018;13(5):e0194768. doi:10.1371/journal.pone.0194768
48. 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
49. Ohmann C, Banzi R, Canham S, et al. Sharing and reuse of individual participant data from clinical trials: principles and recommendations. *BMJ Open*. 2017;7(12):e018647. doi:10.1136/bmjopen-2017-018647
50. Strom BL, Buyse M, Hughes J, Knoppers BM. Data sharing, year 1: access to data from industry-sponsored clinical trials. *N Engl J Med*. 2014;371(22):2052-2054. doi:10.1056/NEJMp1411794
51. Anderson MS, Ronning EA, Devries R, Martinson BC. Extending the mertonian norms: scientists' subscription to norms of research. *J Higher Educ*. 2010;81(3):366-393. doi:10.1080/00221546.2010.11779057
52. Mehra MR, Ruschitzka F, Patel AN. Retraction: hydroxychloroquine or chloroquine with or without a macrolide for treatment of COVID-19: a multinational registry analysis. *Lancet*. 2020;395(10240):1820. doi:10.1016/S0140-6736(20)31324-6
53. Mehra MR, Desai SS, Kuy S, Henry TD, Patel AN. Retraction: cardiovascular disease, drug therapy, and mortality in Covid-19. *N Engl J Med*. 2020;382(26):2582. doi:10.1056/NEJMoa2007621
54. The Editors of the Lancet Group. Learning from a retraction. *Lancet*. 2020;396(10257):1056. doi:10.1016/S0140-6736(20)31958-9
55. Barton CM, Alberti M, Ames D, et al. Call for transparency of COVID-19 models. *Science*. 2020;368(6490):482-483.
56. Ewers M, Ioannidis JPA, Plesnila N. Access to data from clinical trials in the COVID-19 crisis: open, flexible, and time-sensitive. *J Clin Epidemiol*. Published online October 14, 2020. doi:10.1016/j.jclinepi.2020.10.008
57. Janiaud P, Axfors C, Van't Hooft J, et al. The worldwide clinical trial research response to the COVID-19 pandemic: the first 100 days. *F1000Res*. 2020;9:1193-1193. doi:10.12688/f1000research.26707.1
58. Naci H, Kesselheim AS, Røttingen J-A, Salanti G, Vandvik PO, Cipriani A. Producing and using timely comparative evidence on drugs: lessons from clinical trials for COVID-19. *BMJ*. 2020;371:m3869. doi:10.1136/bmj.m3869
59. Dean NE, Gsell P-S, Brookmeyer R, et al. Creating a framework for conducting randomized clinical trials during disease outbreaks. *N Engl J Med*. 2020;382(14):1366-1369. doi:10.1056/NEJMs1905390

#### SUPPLEMENT.

**eTable.** Description of Clinical Trial Data Repositories

**eAppendix 1.** Examples of Data Sharing Statements

**eFigure.** Flow Chart of Article Eligibility

**eAppendix 2.** Codebook

**eAppendix 3.** Inclusion Criteria, Search Strategy, and Data Collection and Analysis