Florian Naudet, Charlotte Sakarovitch, Perrine Janiaud, Ioana Cristea, Daniele Fanelli, David Moher and John P. A. Ioannidis

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

Article (Published version) (Refereed)

Original citation:
Naudet, Florian and Sakarovitch, Charlotte and Janiaud, Perrine and Cristea, Ioana and Fanelli, Daniele and Moher, David and Ioannidis, John P. A. (2018) 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, 360 (k400). ISSN 1756-1833
DOI: 10.1136/bmj.k400

© 2018 British Medical Journal Publishing Group
CC-BY-NC 4.0

This version available at: http://eprints.lse.ac.uk/87729/
Available in LSE Research Online: May 2018

LSE has developed LSE Research Online so that users may access research output of the School. Copyright © and Moral Rights for the papers on this site are retained by the individual authors and/or other copyright owners. Users may download and/or print one copy of any article(s) in LSE Research Online to facilitate their private study or for non-commercial research. You may not engage in further distribution of the material or use it for any profit-making activities or any commercial gain. You may freely distribute the URL (http://eprints.lse.ac.uk) of the LSE Research Online website.
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*

Florian Naudet,1 Charlotte Sakarovitch,2 Perrine Janiaud,1 Ioana Cristea,1,3 Daniele Fanelli,1,4 David Moher,1,5 John P A Ioannidis1,6

**ABSTRACT**

**OBJECTIVES**

To explore the effectiveness of data sharing by randomized controlled trials (RCTs) in journals with a full data sharing policy and to describe potential difficulties encountered in the process of performing reanalyses of the primary outcomes.

**DESIGN**

Survey of published RCTs.

**SETTING**

PubMed/Medline.

**ELIGIBILITY CRITERIA**

RCTs that had been submitted and published by *The BMJ* and *PLOS Medicine* subsequent to the adoption of data sharing policies by these journals.

**MAIN OUTCOME MEASURE**

The primary outcome was data availability, defined as the eventual receipt of complete data with clear labelling. Primary outcomes were reanalyzed to assess to what extent studies were reproduced. Difficulties encountered were described.

**RESULTS**

37 RCTs (21 from *The BMJ* and 16 from *PLOS Medicine*) published between 2013 and 2016 met the eligibility criteria. 17/37 (46%, 95% confidence interval 30% to 62%) satisfied the definition of data availability and 14 of the 17 (82%, 59% to 94%) were fully reproduced on all their primary outcomes. Of the remaining RCTs, errors were identified in two but reached similar conclusions and one paper did not provide enough information in the Methods section to reproduce the analyses. Difficulties identified included problems in contacting corresponding authors and lack of resources on their behalf in preparing the datasets. In addition, there was a range of different data sharing practices across study groups.

**CONCLUSIONS**

Data availability was not optimal in two journals with a strong policy for data sharing. When investigators shared data, most reanalyses largely reproduced the original results. Data sharing practices need to become more widespread and streamlined to allow meaningful reanalyses and reuse of data.

**TRIAL REGISTRATION**

Open Science Framework osf.io/c4zke.

**Introduction**

Patients, medical practitioners, and health policy analysts are more confident when the results and conclusions of scientific studies can be verified. For a long time, however, verifying the results of clinical trials was not possible, because of the unavailability of the data on which the conclusions were based. Data sharing practices are expected to overcome this problem and to allow for optimal use of data collected in trials: the value of medical research that can inform clinical practice increases with greater transparency and the opportunity for external researchers to reanalyze, synthesize, or build on previous data.

In 2016 the International Committee of Medical Journal Editors (ICMJE) published an editorial1 stating that “it is an ethical obligation to responsibly share data generated by interventional clinical trials because participants have put themselves at risk.” The ICMJE proposed to require that deidentified individual patient data (IPD) are made publicly available no later than six months after publication of the trial results. This proposal triggered debate.2-7 In June 2017, the ICMJE stepped back from its proposal. The new requirements do not mandate data sharing but only a data sharing plan to be included in each paper (and prespecified in study registration).8

Because of this trend toward a new norm where data sharing for randomized controlled trials (RCTs) becomes a standard, it seems important to assess how accessible the data are in journals with existing data sharing policies. Two leading general medical journals, *The BMJ*9 and *PLOS Medicine*,10 already have a policy expressly requiring data sharing as a condition for publication of clinical trials: data sharing became a requirement after January 2013 for RCTs on drugs and devices9 and July 2015 for all therapeutics10.
We explored the effectiveness of RCT data sharing in both journals in terms of data availability, feasibility, and accuracy of reanalyses and describe potential difficulties encountered in the process of performing reanalyses of the primary outcomes. We focused on RCTs because they are considered to represent high quality evidence and because availability of the data is crucial in the evaluation of health interventions. RCTs represent the most firmly codified methodology, which also allows data to be most easily analyzed. Moreover, RCTs have been the focus of transparency and data sharing initiatives, owing to the importance of primary data availability in the evaluation of therapeutics (eg, for IPD meta-analyses).

**Methods**

The methods were specified in advance. They were documented in a protocol submitted for review on 12 November 2016 and subsequently registered with the OpenScienceFramework on 15 November 2016 (https://osf.io/u6hcv/register/565fb3678c5e4a66b5582f67).

**Eligibility criteria**

We surveyed publications of RCTs, including cluster trials and crossover studies, non-inferiority designs, and superiority designs, that had been submitted and published by The BMJ and PLOS Medicine subsequent to the adoption of data sharing policies by these journals.

**Search strategy and study selection**

We identified eligible studies from PubMed/Medline. For The BMJ we used the search strategy: “BMJ”[jour] AND (“2013/01/01”[PDAT]: “2017/01/01”[PDAT]) AND Randomized Controlled Trial[ptyp]. For PLOS Medicine we used: “PLoS Med”[jour] AND (“2014/03/01”[PDAT]: “2017/01/01”[PDAT]) AND Randomized Controlled Trial[ptyp].

Two reviewers (FN and PJ) performed the eligibility assessment independently. Disagreements were resolved by consensus or in consultation with a third reviewer (JPAI or DM). More specifically, the eligibility assessment was based on the date of submission, not on the date of publication. When these dates were not available we contacted the journal editors for them.

**Data extraction and datasets retrieval**

A data extraction sheet was developed. For each included study we extracted information on study characteristics (country of corresponding author, design, sample size, medical specialty and disease, and funding), type of intervention (drug, device, other), and procedure to gather the data. Two authors (FN and PJ) independently extracted the data from the included studies. Disagreements were resolved by consensus or in consultation with a third reviewer (JPAI). One reviewer (FN) was in charge of retrieving the IPD for all included studies by following the instructions found in the data sharing statement of the included studies. More specifically, when data were available on request, we sent a standardized email (https://osf.io/h9cas/). Initial emails were sent from a professional email address (fnaudet@stanford.edu), and three additional reminders were sent to each author two or three weeks apart, in case of non-response.

**Data availability**

Our primary outcome was data availability, defined as the eventual receipt of data presented with sufficient information to reproduce the analysis of the primary outcomes of the included RCTs (ie, complete data with clear labelling). Additional information was collected on type of data sharing (request by email, request using a specific website, request using a specific register, available on a public register, other), time for collecting the data (in days, time between first attempt to success of getting a database), deidentification of data (concerning name, birthdate, and address), type of data shared (from case report forms to directly analyzable datasets), sharing of analysis code, and reasons for non-availability in case data were not shared.

**Reproducibility**

When data were available, a single researcher (FN) carried out a reanalysis of the trial. For each study, analyses were repeated exactly as described in the published report of the study. Whenever insufficient details about the analysis was provided in the study report, we sought clarifications from the trial investigators. We considered only analyses concerning the primary outcome (or outcomes, if multiple primary outcomes existed) of each trial. Any discrepancy between results obtained in the reanalysis and those reported in the publication was examined in consultation with a statistician (CS). This examination aimed to determine if, based on both quantitative (effect size, P values) and qualitative (clinical judgment) consideration, the discrepant results of the reanalysis entailed a different conclusion from the one reported in the original publication. Any disagreement or uncertainty over such conclusions was resolved by consulting a third coauthor with expertise in both clinical medicine and statistical methodology (JPAI).

If, after this assessment process, it was determined that the results (and eventually conclusions) were still not reproduced, CS independently reanalyzed the data to confirm such a conclusion. Once the “not reproduced” status of a publication was confirmed, FN contacted the authors of the study to discuss the source of the discrepancy. After this assessment procedure, we classified studies into four categories: fully reproduced, not fully reproduced but same conclusion, not reproduced and different conclusion, and not reproduced (or partially reproduced) because of missing information.

**Difficulties in getting and using data or code and performing reanalyses**

We noted whether the sharing of data or analytical code, or both, required clarifications for which
additional queries had to be presented to the authors to obtain the relevant information, clarify labels or use, or both, and reproduce the original analysis of the primary outcomes. A catalogue of these queries was created and we grouped similar clarifications for descriptive purposes to generate a list of some common challenges and to help tackle these challenges preemptively in future published trials.

**Statistical analyses**

We computed percentages of data sharing or reproducibility with 95% confidence intervals based on binomial approximation or on Wilson score method without continuity correction if necessary. For the purposes of registration, we hypothesized that if these data sharing policies were effective they would lead to more than 80% of studies sharing their data (ie, the lower boundary of the confidence interval had to be more than 80%). High rates of data sharing resulting from full data sharing policies should be expected. On the basis of experience of explicit data sharing policies in *Psychological Science,* however, we knew that a rate of 100% was not realistic and judged that an 80% rate could be a desirable outcome.

When data were available, one researcher (FN) performed reanalyses using the open source statistical software R (R Development Core Team), and the senior statistician (CS) used SAS (SAS Institute). In addition, when authors shared their codes (in R, SAS, STATA (StataCorp 2017) or other), these were checked and used. Estimates of effect sizes, 95% confidence intervals, and P values were obtained for each reanalysis.

**Changes from the initial protocol**

Initially we planned to include all studies published after the data sharing policies were in place. Nevertheless, some authors who we contacted suggested that their studies were not eligible because the papers were submitted to the journal before the policy. We contacted editors who confirmed that policies applied for papers submitted (and not published) after the policy was adopted. In accordance, and to avoid any underestimation of data sharing rates, we changed our selection criteria to “studies submitted and published after the policy was adopted.” For a few additional studies submitted before the policy, data were collected and reanalyzed but only described in the web appendix.

For the reanalysis we initially planned to consider non-reproducibility as a disagreement between reanalyzed results and results reported in the original publication by more than 2% in the point estimate or 95% confidence interval. After reanalysis of a couple of studies we believed that such a definition was sometimes meaningless. Interpreting a RCT involves clinical expertise and cannot be reduced to solely quantitative factors. Accordingly, we changed our definition and provided a detailed description of reanalyses and published results, mentioning the nature of effect size and the type of outcome considered.

**Patient involvement**

We had no established contacts with specific patient groups who might be involved in this project. No patients were involved in setting the research question or the outcome measures, nor were they involved in the design and implementation of the study. There are no plans to involve patients in the dissemination of results, nor will we disseminate results directly to patients.

**Results**

**Characteristics of included studies**

Figure 1 shows the study selection process. The searches done on 12 November 2016 resulted in 159 citations. Of these, 134 full texts were considered for eligibility. Thirty seven RCTs (21 from *The BMJ* and 16 from *PLOS Medicine*) published between 2013 and 2016 met our eligibility criteria. Table 1 presents

---

**Table 1: Summary of reanalysis**

<table>
<thead>
<tr>
<th>Reanalysis Result</th>
<th>The BMJ</th>
<th>PLOS Medicine</th>
</tr>
</thead>
<tbody>
<tr>
<td>Analyses fully reproduced</td>
<td>(n=14)</td>
<td>(n=7)</td>
</tr>
<tr>
<td>Analyses not fully reproduced because of missing information</td>
<td>(n=1)</td>
<td>(n=1)</td>
</tr>
<tr>
<td>Analyses not fully reproduced but same conclusion</td>
<td>(n=2)</td>
<td>(n=1)</td>
</tr>
</tbody>
</table>

---

**Fig 1 | Study flow diagram**

Records identified through database searching (n=159):
The BMJ (n=120) PLOS Medicine (n=39)

Records excluded based on title and abstract (n=25):
The BMJ (non-randomised controlled trials) (n=20)
PLOS Medicine (non-randomised controlled trials) (n=5)

Full text considered for eligibility (n=134):
The BMJ (n=100) PLOS Medicine (n=14)

Records excluded based on full text (n=72):
The BMJ (n=68):
No policy (n=55)
Reanalyses (n=2)
Secondary analyses (n=11)
PLOS Medicine (n=4):
Secondary analyses (n=4)

Full text meeting inclusion criteria published after the policy (n=62):
The BMJ (n=12)
PLOS Medicine (n=30)

Record excluded because submitted before the policy (n=25):
The BMJ (n=11)
PLOS Medicine (n=14)

Full text meeting inclusion criteria submitted after the policy (n=37):
The BMJ (n=21)
PLOS Medicine (n=16)

Data not available (n=20):
The BMJ (n=13)
PLOS Medicine (n=7)

Data available (n=17):
The BMJ (n=8)
PLOS Medicine (n=9)

Analyses fully reproduced (n=14):
The BMJ (n=7)
PLOS Medicine (n=7)

Analyses not reproduced because of missing information (n=1):
PLOS Medicine (n=1)

Analyses not fully reproduced but same conclusion (n=2):
The BMJ (n=1)
PLOS Medicine (n=1)
The characteristics of these studies. These RCTs had a median sample size of 432 participants (interquartile range 213-1070), had no industry funding in 26 cases (70%), and were led by teams from Europe in 25 cases (67%). Twenty RCTs (54%) evaluated pharmacological interventions, 9 (24%) complex interventions (eg, psychotherapeutic program), and 8 (22%) devices.

Data availability
We were able to access data for 19 out of 37 studies (51%). Among these 19 studies, the median number of days for collecting the data was 4 (range 0-191). Two of these studies, however, did not provide sufficient information within the dataset to enable direct reanalysis (eg, had unclear labels). Therefore 17 studies satisfied our definition of data availability. The rate of data availability was 46% (95% confidence interval 30% to 62%).

Data were in principle available for two additional studies not included in the previous count and both authored by the same research team. However, the authors asked us to cover the financial costs of preparing the data for sharing (£607; $857; €694). Since other teams shared the data for free, we considered that it would not have been fair to pay some and not others for similar work in the context of our project and so we classified these two studies as not sharing data. For a third study, the authors were in correspondence with us and discussing conditions for sharing data, but we did not receive that data by the time our data collection process was determined to be over (seven months). If these three studies were included, the proportion of data sharing would be 54% (95% confidence interval 38% to 70%).

For the remaining 15 studies classified as not sharing data, reasons for non-availability were: no answer to...
the different emails (n=7), no answer after an initial agreement (n=2), and refusal to share data (n=6). Explanations for refusal to share data included lack of endorsement of the objectives of our study (n=1), personal reasons (e.g., sick leave, n=2), restrictions owing to an embargo on data sharing (n=1), and no specific reason offered (n=2). The existence of possible privacy concerns was never put forward as a reason for not sharing data.

Among the 19 studies sharing some data (analyzable datasets and non-analyzable datasets), 16 (84%) datasets were totally deidentified. Birthdates were found in three datasets and geographical information (country and postcode) in one of these three. Most datasets were data ready for analysis (n=17), whereas two required some additional processing before the analysis could be repeated. In these two cases, such processing was difficult to implement (even with the code being available) and the authors were contacted to share analyzable data. Statistical analysis code was available for seven studies (including two that were obtained after a second specific request).

Reproducibility
Among the 17 studies16–32 providing sufficient data for reanalysis of their primary outcomes, 14 (82%, 95% confidence interval 59% to 94%) studies were fully reproduced on all their primary outcomes. One of the 17 studies did not provide enough information in the Methods section for the analyses to be reproduced (specifically, the methods used for adjustment were unclear). We contacted the authors of this study to obtain clarifications but received no reply. Of the remaining 16 studies, we reanalyzed 47 different primary analyses. Two of these studies were considered not fully reproduced. For one study, we identified an error in the statistical code and for the other we found slightly different numerical values for the effect sizes measured as well as slight differences in numbers of patients included in the analyses (a difference of one patient in one of four analyses). Nevertheless, similar conclusions were reached in both cases (these two studies were categorized as not fully reproduced but reaching the same conclusion). Therefore, we found no results contradicting the initial publication, neither in terms of magnitude of the effect (table 2) nor in terms of statistical significance of the finding (fig 2).

We retrieved the data of three additional studies, published in The BMJ after its data sharing policy was in place (but submitted before the policy). Although these studies were ineligible for our main analysis, reanalyses were performed and also reached the same conclusions as the initial study (see supplementary e-Table 1).

Difficulties in getting and using data or code and performing reanalyses
Based on our correspondence with authors, we identified several difficulties in getting the data. A common concern pertained to the costs of the data sharing process—for example, the costs of preparing the data or translating the database from one language to another. Some authors wondered whether their team or our team should assume these costs. In addition, some of the authors balanced these additional costs with their perceived benefits of sharing data for the purpose of this study and seemed to rather value data sharing for the purpose of a meta-analysis than for reanalyses, possibly because of the risk acknowledged by one investigator we contacted about “naming and shaming” individual studies or investigators.

Getting prepared and preplanning for data sharing still seems to be a challenge for many trial groups; data sharing proved to be novel for some authors who were unsure how to proceed. Indeed, there was considerable heterogeneity between different procedures to share data: provided in an open repository (n=5), downloadable on a secured website (n=1) after registration, included as appendix of the published paper (n=3), or sent by email (n=10). On three occasions, we signed a data sharing request or agreement. In these agreements, the sponsor and recipient parties specified the terms that bound them in the data sharing process (e.g., concerning the use of data, intellectual property, etc). In addition, typically there was no standard in what type of data were shared (at what level of cleaning and processing). In one case, authors mentioned explicitly that they followed standardized guidelines33 to prepare the dataset.

Some analyses were complex and it was sometimes challenging to reanalyze data from specific designs or when unusual measures were used (e.g., relative change in percentage). Obtaining more information about the analysis by contacting authors was necessary for 6 of 17 studies to replicate the findings. In one case, specific exploration of the code revealed that in a survival analysis, authors treated repeated events in the same patient (multiple events) as distinct observations. This was in disagreement with the methods section describing usual survival analysis (taking into account only the first event for each patient). However, alternative analyses did not contradict the published results.

Three databases did not provide sufficient information to reproduce the analyses. Missing data concerned variables used for adjustment, definition of the analysis population, and randomization groups. Communication with authors was therefore necessary and was fruitful in one of these three cases.

Discussion
In two prominent medical journals with a strong data sharing rate is much higher than the average rate of published articles the original investigators shared their data with sufficient information to enable reanalyses. This rate was less than the 80% boundary that we prespecified as an acceptable threshold for papers submitted under a policy that makes data sharing an explicit condition for publication. However, despite being lower than might be desirable, a 46% data sharing rate is much higher than the average rate.
<table>
<thead>
<tr>
<th>Study, treatment comparison; condition</th>
<th>Primary outcome measures</th>
<th>Effect size (type)</th>
<th>Effect size: published paper</th>
<th>P value</th>
<th>Effect size: reanalysis</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Luedtke 2015*</td>
<td>transcranial direct current stimulation (TDCs) versus sham TDCs; chronic low back pain</td>
<td>Visual analogue scale, time 1</td>
<td>MD (99% CI) 1</td>
<td>(−8.69 to 6.30)</td>
<td>0.68</td>
<td>1.42 (−0.60 to 8.90)</td>
</tr>
<tr>
<td></td>
<td>Oswestry disability index, time 1</td>
<td>MD (99% CI) 1</td>
<td>(−1.73 to 1.98)</td>
<td>0.86</td>
<td>−0.12 (−1.99 to 1.75)</td>
<td>0.87</td>
</tr>
<tr>
<td>Cohen 2015†</td>
<td>epidural steroid injections versus gabapentin; lumbosacral radicular pain</td>
<td>Average leg pain, time 1</td>
<td>MD (99% CI) 0.4</td>
<td>(−0.3 to 1.2)</td>
<td>0.25</td>
<td>0.4 (−0.3 to 1.2)</td>
</tr>
<tr>
<td></td>
<td>Oswestry disability index, time 1</td>
<td>MD (99% CI) 0.3</td>
<td>(−0.5 to 1.2)</td>
<td>0.43</td>
<td>0.3 (−0.5 to 1.2)</td>
<td>0.45</td>
</tr>
<tr>
<td>Hylte Sorensen 2015*</td>
<td>near infrared spectroscopy versus usual care; preterm infants</td>
<td>Burden of hypoxia and hyperoxia</td>
<td>RC% (95% CI) −58</td>
<td>(−35 to −74)</td>
<td>&lt;0.001</td>
<td>−58 (−36 to −72)</td>
</tr>
<tr>
<td>Bousema 2016†</td>
<td>hotspot targeted interventions versus usual interventions; malaria</td>
<td>nPCR parasite prevalence, time 1, zone 1</td>
<td>MD (95% CI) 4.2</td>
<td>(−5.1 to 13.8)</td>
<td>0.326; 0.265</td>
<td>4.3 (−5.3 to 13.9)</td>
</tr>
<tr>
<td></td>
<td>nPCR parasite prevalence, time 1, zone 2</td>
<td>MD (95% CI) 3.6</td>
<td>(−2.6 to 9.7)</td>
<td>0.219; 0.216</td>
<td>3.6 (−3.3 to 0.1)</td>
<td>0.339; 0.339</td>
</tr>
<tr>
<td></td>
<td>nPCR parasite prevalence, time 2, zone 2</td>
<td>MD (95% CI) 1.0</td>
<td>(−7.0 to 9.1)</td>
<td>0.775; 0.713</td>
<td>1.0 (−7.6 to 9.7)</td>
<td>0.778; 0.858</td>
</tr>
<tr>
<td>Gysin-Maillart 2016‡</td>
<td>brief therapy versus usual care; suicide attempt</td>
<td>Suicide attempt</td>
<td>HR (95% CI) 0.17</td>
<td>(0.07 to 0.46)</td>
<td>&lt;0.001</td>
<td>0.17 (0.06 to 0.49)</td>
</tr>
<tr>
<td>Lombard 2016*</td>
<td>low intensity intervention versus general health intervention; obesity</td>
<td>Change in weight (unadjusted)</td>
<td>MD (95% CI) −0.92</td>
<td>(−1.67 to −0.16)</td>
<td>&lt;0.001</td>
<td>−0.92 (−1.66 to −0.18)</td>
</tr>
<tr>
<td>Polyak 2016*</td>
<td>cotrimoxazole prophylaxis cessation versus continuation; HIV</td>
<td>Malaria, pneumonia, diarrhea, or mortality (ITT)</td>
<td>RR (95% CI) 2.27</td>
<td>(1.51 to 3.39)</td>
<td>&lt;0.001</td>
<td>2.27 (1.51 to 3.39)</td>
</tr>
<tr>
<td>Robinson 2016†</td>
<td>blood stage plus liver stage drugs or blood stage drugs only; malaria</td>
<td>Plasmodium vivax infection by qPCR, time 1</td>
<td>HR (95% CI) 0.18</td>
<td>(0.14 to 0.25)</td>
<td>&lt;0.001</td>
<td>0.18 (0.14 to 0.25)</td>
</tr>
<tr>
<td>Gysin-Maillart 2016‡</td>
<td>brief therapy versus usual care; suicide attempt</td>
<td>Daily step count</td>
<td>MD (95% CI) 1037</td>
<td>(513 to 1560)</td>
<td>&lt;0.001</td>
<td>1037 (513 to 1560)</td>
</tr>
<tr>
<td>Adron 2016*</td>
<td>low dose betahistine versus placebo; Meniere’s disease</td>
<td>Vertigo attack</td>
<td>RaR (95% CI) 1.036</td>
<td>(0.942 to 1.140)</td>
<td>0.759</td>
<td>1.036 (0.942 to 1.133)</td>
</tr>
<tr>
<td>Selak 2014‡</td>
<td>fixed dose combination treatment versus usual care; patients at high risk of cardiovascular disease in primary care</td>
<td>Use of antiplatelet, statin, and 2 blood pressure lowering drugs</td>
<td>RR (95% CI) 1.75</td>
<td>(1.52 to 2.03)</td>
<td>&lt;0.001</td>
<td>1.75 (1.52 to 2.03)</td>
</tr>
<tr>
<td>Chesterton 2013*</td>
<td>transcutaneous electrical nerve stimulation versus usual care; tennis elbow</td>
<td>Change in elbow pain intensity</td>
<td>MD (95% CI) −0.33</td>
<td>(−0.96 to 0.31)</td>
<td>0.31</td>
<td>−0.33 (−1.01 to 0.34)</td>
</tr>
<tr>
<td>Atukunda 2014‡</td>
<td>sublingual misoprostol versus intramuscular oxytocin; prevention of postpartum haemorrhage</td>
<td>Blood loss &gt;500 mL</td>
<td>RR (95% CI) 1.64</td>
<td>(1.32 to 2.05)</td>
<td>&lt;0.001</td>
<td>1.65 (1.32 to 2.05)</td>
</tr>
<tr>
<td>Paul 2015*</td>
<td>trimethoprim-sulfamethoxazole versus vancomycin; severe infections caused by meticillin resistant Staphylococcus aureus</td>
<td>Treatment failure (ITT)</td>
<td>RR (95% CI) 1.38</td>
<td>(0.96 to 1.99)</td>
<td>0.08</td>
<td>1.38 (0.96 to 1.99)</td>
</tr>
<tr>
<td></td>
<td>Treatment failure (PP)</td>
<td>RR (95% CI) 1.24</td>
<td>(0.82 to 1.89)</td>
<td>0.36</td>
<td>1.24 (0.82 to 1.89)</td>
<td>0.36</td>
</tr>
<tr>
<td></td>
<td>All cause mortality (ITT)</td>
<td>RR (95% CI) 1.27</td>
<td>(0.65 to 2.45)</td>
<td>0.57</td>
<td>1.27 (0.65 to 2.45)</td>
<td>0.57</td>
</tr>
<tr>
<td></td>
<td>All cause mortality (PP)</td>
<td>RR (95% CI) 1.05</td>
<td>(0.67 to 1.62)</td>
<td>0.1</td>
<td>1.05 (0.67 to 1.62)</td>
<td>0.1</td>
</tr>
<tr>
<td>Omerod 2015‡</td>
<td>ciclosporin versus prednisolone; pyoderma gangrenosum</td>
<td>Speed of healing over six weeks (cm²/day)</td>
<td>MD (95% CI) 0.003</td>
<td>(−0.20 to 0.21)</td>
<td>0.97</td>
<td>0.003 (−0.20 to 0.21)</td>
</tr>
<tr>
<td>Hanson 2015‡</td>
<td>home based counselling strategy versus usual care; neonatal care</td>
<td>Neonatal mortality</td>
<td>OR (95% CI) 1.0</td>
<td>(0.9 to 1.2)</td>
<td>0.339</td>
<td>1.0 (0.9 to 1.2)</td>
</tr>
</tbody>
</table>

nPCR = nested polymerase chain reaction; qPCR = quantitative real time PCR; MD = mean differences; RC% = relative change in percentage; HR = hazard Ratio; RaR = rate ratio; RR = relative risk; OR = odds ratio; ITT = intention to treat; PP = per protocol.
*Study judged as not fully reproduced but reaching same conclusion because of differences in numerical estimates; for one of these analyses current authors found a difference in number of patients analyzed (one patient more in reanalysis).
†Study judged as not fully reproduced but reaching same conclusion because initial analysis did not take into account repeated nature of data. This was corrected in reanalysis without major differences in estimates.
‡Study judged as not fully reproduced but reaching same conclusion because initial analysis did not take into account repeated nature of data. This was corrected in reanalysis without major differences in estimates.
of biomedical literature at large, in which data sharing is almost non-existent (with few exceptions in some specific disciplines, such as genetics). Moreover, our analyses focused on publications that were submitted directly after the implementation of new data sharing policies, which might be expected to have practical and cultural barriers to their full implementation. Indeed, our correspondence with the authors helped identify several practical difficulties connected to data sharing, including difficulties in contacting corresponding authors, and lack of time and financial resources on their behalf in preparing the datasets for us. In addition, we found a wide variety of data sharing practices between study groups (ie, regarding the type of data that can be shared and the procedures that are necessary to follow to get the data). Data sharing practices could evolve in the future to deal with these barriers to data sharing (table 3).

For all results that we were able to reanalyze, we reached similar conclusions (despite occasional slight differences in the numerical estimations) to those reported in the original publication, and this result that at least the available data shared do correspond closely to the reported results is reassuring. Of course, there is a large amount of diversity on what exactly “raw data” mean and they can involve various transformations (from the case report forms to coded and analyzable data). Here, we relied on late stage, coded, and cleaned data and therefore the potential for leading to a different conclusion was probably small. Data processing, coding, cleaning, and recategorization of events can have a substantial impact on the results in some trials. For example, SmithKline Beecham’s Study 329 was a well known study on paroxetine in adolescent depression, presenting the drug as safe and effective, whereas a reanalysis starting from the case report forms found a lack of efficacy and some serious safety issues.

**Strengths and weaknesses of this study**

Some leading general medical journals—New England Journal of Medicine, Lancet, JAMA, and JAMA Internal Medicine, have put in place new guidelines for data sharing. Some leading journals may have more efficiently adopted some clear rules and it is possible to see that adherence to those rules is increasing but with large differences in the quality of data sharing.

**Table 3 | Some identified challenges (and suggestions) for data sharing and reanalyses**

<table>
<thead>
<tr>
<th>Identified problems</th>
<th>Suggested solutions for various stakeholders</th>
</tr>
</thead>
<tbody>
<tr>
<td>Data sharing policies leaves responsibility and burden to researchers, to obtain all necessary resources (time, money, technical and organizational tools and services, ethical and legal compliance, etc)</td>
<td>Patients and clinicians should help to develop awareness of a common ownership of the data, intended as common responsibility also in providing all necessary resources to make data sharable for effective and ethical use</td>
</tr>
<tr>
<td>Getting prepared and preplanning for data sharing is still in progress in trials units. There is considerable heterogeneity between different procedures to share data and types of data that are shared</td>
<td>Researchers should pre-emptively address and seek funding for data sharing Funders should allow investigators to use funds towards data sharing Academic institutions should reward data sharing activities through promotion and tenure</td>
</tr>
<tr>
<td>Data sharing on request leaves researchers exclusive discretion on decision about whether to share their own data to other research groups, for which objectives, and under which terms and conditions</td>
<td>All stakeholders should adopt some clear and homogeneous rules (eg, guidelines) for best practices in data sharing Clinical trials groups should develop comprehensive educational outreach about data sharing Institutional review boards should systematically address data sharing issues</td>
</tr>
<tr>
<td>There could be some difficulties in contacting corresponding authors, limiting data sharing on request</td>
<td>Researchers and editors should favor data deposition when it is ethically possible</td>
</tr>
<tr>
<td>Some identifying information such as date of birth can be found in some databases</td>
<td>Researchers must ensure that databases are deidentified before sharing Institutional review board should provide guidance on the requested level of deidentification for each individual study. It included seeking consent for sharing IPD from trial participants</td>
</tr>
<tr>
<td>Shared databases need to be effectively sharable (eg, complete, homogeneous), including meta-data (eg, descriptive labels, description of pre-analysis processing tools, methods of analysis)</td>
<td>Clinical trials groups should develop comprehensive educational outreach about data sharing Researchers should provide details concerning the detailed labels used in the table and analytical code</td>
</tr>
<tr>
<td>Reproducible research practices are not limited to sharing data, materials, and code. Complete reporting of methods and statistical analyses are also relevant</td>
<td>Clinical trials groups should develop comprehensive educational outreach about data sharing Researchers should also share their detailed statistical analysis plan Reporting guidelines should emphasize best practices in data sharing about computational reproducibility</td>
</tr>
</tbody>
</table>
Medicine—have had no specific policy for data sharing in RCTs until recently. Annals of Internal Medicine has encouraged (but not demanded) data sharing since 2007.36 BMC Medicine adopted a similar policy in 2015. The BMJ and PLOS Medicine have adopted stronger policies, beyond the ICMJE policy, that mandate data sharing for RCTs. Our survey of RCTs published in these two journals might therefore give a taste of the impact and caveats of such full policies. However, care should be taken to not generalize these results to other journals. First, we had a selected sample of studies. Our sample included studies (mostly from Europe) that are larger and less likely to be funded by the industry than the average published RCT in the medical literature.39 Several RCTs, especially those published in PLOS Medicine, were cluster randomized studies and many explored infectious disease or important public health issues, characteristics that are not common in RCTs overall. Some public funders (or charities) involved in their funding have already open access policies, such as the Bill and Melinda Gates foundation or the UK National Institute for Health Research.

These reasons also explain why we did not compare journals. Qualitatively they do not publish the same kind of RCTs, and quantitatively The BMJ and PLOS Medicine publish few RCTs compared with other leading journals, such as the New England Journal of Medicine and Lancet. Any comparisons might have been subject to confounding. Similarly, we believed that before and after studies at The BMJ and PLOS Medicine might have been subject to historical bias and confounding factors since such policies might have changed the profile of submitting authors: researchers with a specific interest in open science and reproducibility might have been attracted and others might have opted for another journal. We think comparative studies will be easier to conduct when all journals adopt data sharing standards. Even with the current proposed ICMJE requirements, The BMJ and PLOS Medicine will be journals with stronger policies. In addition, The BMJ and PLOS Medicine are both major journals with many resources, such as in-depth discussion of papers in editorial meetings and statistical peer review. Our findings concerning reproducibility might not apply to smaller journals with more limited resources. We cannot generalize this finding to studies that did not share their data: authors who are confident in their results might more readily agree to share their data and this may lead to overestimation of reproducibility. In addition, we might have missed a few references based on primary outcomes, whereas secondary and safety outcomes may be more problematic in their reproduction. These points need to be explored in further studies.

Strengths and weaknesses in relation to other studies

In a precedent survey of 160 randomly sampled research articles in The BMJ from 2009 to 2015, excluding meta-analyses and systematic reviews,40 the authors found that only 5% shared their datasets. Nevertheless, this survey assessed data sharing among all studies with original raw data, whereas The BMJ/data sharing policy specifically applied to clinical trial data. When considering clinical trials bound by The BMJ data sharing policy (n=21), the percentage shared was 24% (95% confidence interval 8% to 47%). Our study identified higher rates of shared datasets in accordance with an increase in the rate of “data shared” for every additional year between 2009 and 2015 found in the previous survey. It is not known whether it results directly from the policy implementation or from a slow and positive cultural change of trialists.

Lack of response from authors was also identified as a caveat in the previous evaluation, which suggested that the wording of The BMJ policy, such as availability on “reasonable request,” might be interpreted in different ways.40 Previous research across PLOS also suggests that data requests by contacting authors might be ineffective sometimes.41 Despite writing a data sharing agreement in their paper, corresponding authors are still free to decline data requests. One could question whether sharing data for the purposes of our project constitutes a “reasonable request.” One might consider that exploring the effectiveness of data sharing policies lies outside the purpose for which the initial trial was done and for which the participants gave consent. A survey found that authors are generally less willing to share their data for the purpose of reanalyses than, for instance, for individual patient data (IPD) meta-analyses.42 Nevertheless, reproducibility checks based on independent reanalysis are perfectly aligned with the primary objective of clinical trials and indeed with patient interests. Conclusions that persist through substantial reanalyses are becoming more credible. To explain and support the interest of our study, we have registered a protocol and have transparently described our intentions in our email, but data sharing rates were still lower than we expected. Active auditing of data sharing policies by journal editors may facilitate the implementation of data sharing.

Concerning reproducibility of clinical trials, an empirical analysis suggests that only a small number of reanalyses of RCTs have been published to date; of these, only a minority was conducted by entirely independent authors. In a previous empirical evaluation of published reanalyses, 35% (13/37) of the reanalyses yielded changes in findings that implied conclusions different from those of the original article as to whether patients should be treated or not or about which patients should be treated.43 In our assessment,
we found no differences in conclusions pertaining to treatment decisions. The difference between the previous empirical evaluation and the current one is probably due to many factors. It is unlikely that published reanalyses in the past would be have been published if they had found the same results and had reached the same conclusions as the original analysis. Therefore, the set of published reanalyses is enriched with discrepant results and conclusions. Moreover, published reanalyses addressed the same question on the same data but using typically different analytical methods. Conversely, we used the same analysis employed by the original paper. In addition, the good computational reproducibility (replication of analyses) we found is only one aspect of reproducibility and should not be over-interpreted. For example, in psychology, numerous laboratories have volunteered to re-run experiments (not solely analyses), with the methods used by the original researchers. Overall, 39 out of 100 studies were considered as successfully replicated. But attempts to replicate psychological studies are often easier to implement than attempts to replicate RCTs, which are often costly and difficult to run. This is especially true for large RCTs.

Meaning of the study: possible explanations and implications for clinicians and policy makers

The ICJME’s requirements adopted in 2017 mandate that a data sharing plan will have to be included in each paper (and prespecified in study registration). Because data sharing among two journals with stronger requirements was not optimal, our results suggest that this not likely to be sufficient to achieve high rates of data sharing. One can imagine that individual authors will agree to write such a statement in line with the promise of publication, but the time and costs involved into such data preparation might lead authors to be reluctant to answer data sharing requests. Interestingly the ICJME also mandates that clinical trials that begin enrolling participants on or after 1 January 2019 must include a data sharing plan in the trial’s registration. An a priori data sharing plan might push more authors to pre-emptively deal with and find funding for sharing, but its eventual impact on sharing is unknown. Funders are well positioned to facilitate data sharing. Some, such as the Wellcome Trust, already allow investigators to use funds towards the charges for open access, which is typically a small fraction of awarded funding. Funders could extend their funding policy and also allow investigators to use a similar small fraction of funding towards enabling data sharing. In addition, patients can help promote a culture of data sharing for clinical trials.

In addition, reproducible research does not solely imply sharing data but also reporting all steps of the statistical analysis. This is one core principle of the CONSORT statement “Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to verify the reported results.” It is none the less sometimes difficult to provide such detailed information in a restricted number of lines (ie, a paper). We suggest that details of the statistical analysis plan have to be provided as well as the detailed labels used in the table, and efficient analytical code sharing is also essential.

We suggest that if journals ask for data and code to be shared, they should ensure that this material is also reviewed by editorial staff (with or without peer reviewers) or at a minimum checked for completeness and basic usability. This could also be positively translated in specific incentives for the paper; for example, many psychology journals use badges as signs of good research practices (including data sharing) adopted by papers. And, beyond incentivizing authors, journals adopting such practices could also be incentivized by a gain in reputation and credibility.

Though ensuring patient privacy and lack of explicit consent for sharing are often cited as major barriers to sharing RCT data (and generally accepted as valid exemptions), none of the investigators we approached mentioned this reason. This could suggest that technical constraints, lack of incentives and dedicated funding, and a general diffidence towards reanalyses might be more germane obstacles to be addressed.

Finally, data sharing practices differed from one team to another. There is thus a need for standardization and for drafting specific guidelines for best practices in data sharing. In the United Kingdom, the Medical Research Council Hubs for Trials Methodology Research has proposed a guidance to facilitate the sharing of IPD for publicly funded clinical trials. Other groups might consider adopting similar guidelines.

Unanswered questions and future research

We recommend some prospective monitoring of data sharing practices to ensure that the new requirements of the ICJME are effective and useful. To this end, it should also be kept in mind that data availability is a surrogate of the expected benefit of having open data. Proving that data sharing rates impacts reuse of the data is a further step. Proving that this reuse might translate into discovery that can change care without generating false positive findings (eg, in series of unreliable a posteriori subgroup analyses) is even more challenging.

This study was made possible through sharing of anonymized individual participant data from the authors of all studies. We thank the authors who were contacted for this study: C Bullen and the National Institute for Health Innovation, S Gilbody, C Hewitt, L Littlewood, C van der Meulen, H van der Aa, S Cohen, M Bicket, T Harris, the STOP GAP study investigators including Kim Thomas, Alan Montgomery, and Nicola Greenlaw, Nottingham University Hospitals NHS Trust, NIHR programme grants for applied research, the Nottingham Clinical Trials Unit, C Polyaek, K Yuhas, C Adron, G Gresen, S Hyttel-Sørensen A Barker, R Morello, K Luedtke, M Paul, D Yahav, L Chesterton, the Arthritis Research UK Primary Care Centre, and C Hanson.

Contributors: FN, DF, and JI conceived and designed the experiments. FN, PJ, CS, IC, and JI performed the experiments. FN and CS analyzed the data. FN, CS, PJ, IC, DF, DM, and JI interpreted the results. FN wrote the first draft of the manuscript. CS, PJ, IC, DF, DM, and JI contributed to the writing of the manuscript. FN, CS, PJ, IC, DF, DM, and JI agreed with the results and conclusions of the manuscript. All authors have read, and confirm that they meet, ICMJE criteria for authorship. All authors had full access to all of the data (including statistical reports and
Competing interests: All authors have completed the ICMJE uniform disclosure form at http://www.icmje.org/coiDisclosureForm.pdf (available on request from the corresponding author) and declare that (1) No authors have support from any company for the submitted work; (2) FN has relationships (travel/accommodations expenses covered/reimbursed) in Servier, BMS, Lundbeck, and Janssen who might have an interest in the work submitted in the previous three years. In the past three years PF received a fellowship/grant from GSK for her PhD as part of a public-private collaboration. CS, IC, DF, DM, and IP4 have no relationship with any company that might have an interest in the work submitted. (3) No author’s spouse, partner, or children have any financial relationships that could be relevant to the submitted work; and (4) none of the authors has any non-financial interests that could be relevant to the submitted work.

Funding: METRICS has been funded by Laura and John Arnold Foundation but there was no direct funding for this study. FN received grants from La Fondation Pierre Deniker, Rennes University Hospital, France (CORRECT. Comité de la Recherche Clinique et Translационelle) and Agence Nationale de la Recherche (ANR). PF is supported by a postdoctoral fellowship from the Laura and John Arnold Foundation, IC was supported by the Laura and John Arnold Foundation and the Romanian National Authority for Scientific Research and Innovation, CNCS–UEFISC, project number PN-II-RU-TE-2014-4-1316 (awarded to IC), and the work of JI is supported from an unrestricted gift from Sue and Bob O’Donnell. The sponsors had no role concerning preparation, review, or approval of the manuscript.

Ethical approval: Not required.

Data sharing: The code is shared on the Open Science Framework (https://osf.io/7psw3/). All datasets that were used are retrievable following the instruction of the original papers.

Transparency: The guarantor (FN) affirms that the manuscript is a honest, accurate, and transparent account of the study brought reporting; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license is an Open Access article distributed in accordance with the instruction of the original papers.

Funding: METRICS has been funded by Laura and John Arnold Foundation but there was no direct funding for this study. FN received grants from La Fondation Pierre Deniker, Rennes University Hospital, France (CORRECT. Comité de la Recherche Clinique et Translационelle) and Agence Nationale de la Recherche (ANR). PF is supported by a postdoctoral fellowship from the Laura and John Arnold Foundation, IC was supported by the Laura and John Arnold Foundation and the Romanian National Authority for Scientific Research and Innovation, CNCS–UEFISC, project number PN-II-RU-TE-2014-4-1316 (awarded to IC), and the work of JI is supported from an unrestricted gift from Sue and Bob O’Donnell. The sponsors had no role concerning preparation, review, or approval of the manuscript.

Ethical approval: Not required.

Data sharing: The code is shared on the Open Science Framework (https://osf.io/7psw3/). All datasets that were used are retrievable following the instruction of the original papers.

Transparency: The guarantor (FN) affirms that the manuscript is a honest, accurate, and transparent account of the study brought reporting; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned (and, if relevant, registered) have been explained.

This is an Open Access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license is an Open Access article distributed in accordance with the instruction of the original papers.


Supplementary information: e-Table 1: results of reanalyses for three ineligible studies published in The BMJ after the policy but submitted before