

RESEARCH

Open Access



Access to unpublished protocols and statistical analysis plans of randomised trials

David Campbell¹, Cassandra McDonald¹, Suzie Cro², Vipul Jairath^{1,3†} and Brennan C. Kahan^{4**†}

Abstract

Background: Access to protocols and statistical analysis plans (SAPs) increases the transparency of randomised trial by allowing readers to identify and interpret unplanned changes to study methods, however they are often not made publicly available. We sought to determine how often study investigators would share unavailable documents upon request.

Methods: We used trials from two previously identified cohorts (*cohort 1*: 101 trials published in high impact factor journals between January and April of 2018; *cohort 2*: 100 trials published in June 2018 in journals indexed in PubMed) to determine whether study investigators would share unavailable protocols/SAPs upon request. We emailed corresponding authors of trials with no publicly available protocol or SAP up to four times.

Results: Overall, 96 of 201 trials (48%) across the two cohorts had no publicly available protocol or SAP (11/101 high-impact cohort, 85/100 PubMed cohort). In total, 8/96 authors (8%) shared some trial documentation (protocol only [$n = 5$]; protocol and SAP [$n = 1$]; excerpt from protocol [$n = 1$]; research ethics application form [$n = 1$]). We received protocols for 6/96 trials (6%), and a SAP for 1/96 trial (1%). Seventy-three authors (76%) did not respond, 7 authors responded (7%) but declined to share a protocol or SAP, and eight email addresses were invalid (8%). A total of 329 emails were sent (an average of 41 emails for every trial which sent documentation). After emailing authors, the total number of trials with an available protocol increased by only 3%, from 52% in to 55%.

Conclusions: Most study investigators did not share their unpublished protocols or SAPs upon direct request. Alternative strategies are needed to increase transparency of randomised trials and ensure access to protocols and SAPs.

Keywords: Randomised trials, Protocols, Statistical analysis plans, Data sharing, Unpublished protocols

Background

Trial protocols and statistical analysis plans (SAPs) provide a blueprint for how a study is to be conducted and analysed. Explicit description of such methods before the trial begins helps to identify and deter inappropriate changes made to study methods that could introduce

bias, such as changes to outcomes [1–16] or statistical methods [4, 10, 17–23], after preliminary examination of study results. Protocols and SAPs require more detail than other sources, such as trial registry websites, particularly around planned statistical methods, and are thus essential to enable critical appraisal of trial results.

However, protocols and SAPs can only help identify and deter inappropriate changes to study methods if they are made publicly available. Although most trials published in high-impact journals have an available protocol (82–95%) [18, 24], most of these are dated from after the trial begins [18, 21, 24], making it impossible to

[†]Vipul Jairath and Brennan C Kahan contributed equally.

*Correspondence: b.kahan@ucl.ac.uk

⁴ MRC Clinical Trials Unit at UCL, London, UK
Full list of author information is available at the end of the article



determine whether changes were made after a preliminary assessment of study data. Furthermore, the above rates are not reflective of the situation at large; only 15% of trials indexed in PubMed had an available protocol [21], and availability of SAPs is even lower with only 46–50% of high-impact trials and 3% of PubMed trials having an available SAP (with most being dated after the trial began) [18, 21, 24].

Given the importance of protocols and SAPs to the understanding and interpretation of trial results, we sought to determine how often study investigators would share these documents upon request

Methods

The protocol for this study is available in the [Supplementary Material](#).

Search strategy and eligibility criteria

In stage 1, we reviewed trials published in high impact factor journals (cohort 1) [18], and trials published in journals indexed in PubMed (cohort 2) [21] to identify how many had publicly available protocols and SAPs (i.e. available alongside the article in [supplementary material](#), previously published in an academic journal, or hosted on a website). Results from this stage have been published. The first cohort included randomised controlled trials published between January and April 2018 from six general high impact medical journals (Annals of Internal Medicine; The BMJ; Journal of the American Medical Association (JAMA); The Lancet; New England Journal of Medicine (NEJM); and PLOS Medicine) [18], and the second cohort included trials published in June 2018 in any journal indexed in PubMed [21]. Full details on search strategies are available in the original publications.

Both cohorts included articles reporting results from a phase 2–4 randomised trial in humans. Articles were excluded if they were a pilot or feasibility study, a phase 1 trial, a non-randomised study, a secondary analysis of previously published trial, had cost-effectiveness as the primary outcome, had more than one trial reported in the article, reported results of an interim analysis, or if the protocol or SAP was not in English.

In stage 2 (reported here), we included trials identified in stage 1 that did not have a publicly available protocol or SAP.

Data collection

We emailed the corresponding author of each trial using the email addresses listed on the published trial manuscript. Authors were emailed up to four times in approximately 3-week intervals until a response was received. If a response was not received after the fourth attempt, the author was considered to be non-responsive.

We originally specified in our protocol that we would only email authors up to three times, but we added a fourth email attempt near the end of our study due to the low number of responses received. We also initially specified that we would email authors in 2-week intervals, although practically this occurred in 2–3 weeks intervals. Emails were sent between October 5, 2020, and December 18, 2020. The content of the email (including subject and text) can be found in the [Supplementary Material](#).

Outcomes

Main outcome measures included the number of authors that sent a protocol, the number that sent a SAP, the number that sent both, and the number of authors that sent both the initial and final versions of their SAPs. We had also intended to include the number of authors sending both the initial and final version of their protocol as a main outcome measure, but most protocols that we received did not contain information on the version number or date created, and thus, we were unable to assess this outcome.

The secondary outcomes pertained to trials for which the authors shared either a protocol or SAP and included whether there were any undisclosed deviations to the planned analysis approach for the trial's primary outcome. The methods used to assess undisclosed deviations were the same as those used previously [18, 21]. We chose to focus on undisclosed deviations to the analysis approach as this information is typically only available in the protocol or SAP, rather than from other sources, such as trial registry websites.

Data extraction

Where authors shared a protocol or SAP, data was extracted from these obtained documents onto a pre-piloted standardised data extraction form by two reviewers independently. Disagreements were resolved by discussion or by a third reviewer where disagreement could not be resolved.

Ethical approval

The study protocol was reviewed by the Office of Human Research Ethics at Western University who confirmed that ethical approval was not required.

Results

Two hundred and one trials were included in stage 1 ($n=101$ high impact cohort; $n=100$ PubMed Cohort) (Fig. 1) [18, 21]. Of these, 96 (48%) did not have a publicly available protocol or SAP and were thus included in stage 2 ($n=11$ high-impact cohort; $n=85$ PubMed cohort). Characteristics of included trials are shown in Table 1.

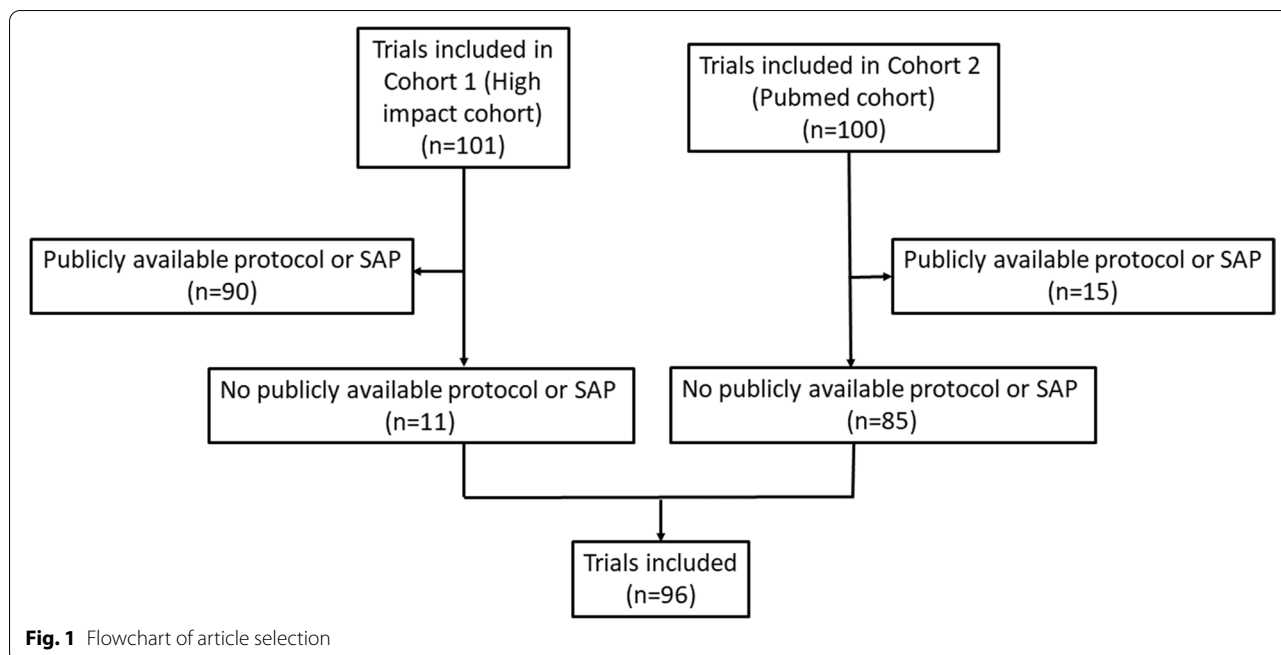


Fig. 1 Flowchart of article selection

Table 1 Baseline characteristics of included trials

Characteristic	High impact cohort (n = 11)	PubMed cohort (n = 85)	Total (n = 96)
Any for profit funding			
Yes	4 (36)	10 (12)	14 (15)
No	7 (64)	44 (52)	51 (53)
Unclear	0 (0)	31 (36)	31 (32)
Type of intervention			
Pharmacologic	5 (45)	39 (46)	44 (46)
Surgical	1 (9)	8 (9)	9 (9)
Psychosocial/behavioural/educational	0 (0)	11 (13)	11 (11)
Other	4 (36)	26 (31)	30 (31)
Multiple types	1 (9)	1 (1)	2 (2)
Cluster trial	1 (9)	2 (2)	3 (3)
Factorial trial	1 (9)	0 (0)	1 (1)
Crossover trial	0 (0)	3 (4)	3 (3)
Non-inferiority trial	4 (36)	4 (5)	8 (8)
No. of treatment arms			
Two	9 (82)	69 (81)	78 (81)
Three or more	2 (18)	16 (19)	18 (19)
Sample size			
Median, IQR	450 (300, 1499)	93 (60, 168)	101 (61, 208)

Data are n (%) unless otherwise indicated

Requesting access to protocols and SAPs

We contacted corresponding authors of all 96 eligible trials via email. Results are shown in Table 2.

Of the 96 authors contacted, 15 (16%) responded, 73 (76%) did not respond, and eight email addresses (8%) were invalid (i.e. returned an error message, such as “user/host unknown”, “access denied”, or “undeliverable”). Of the 15 authors who responded, almost half (47%) required more than one email, and the mean number of days to first response was 17.1 (SD 17.4) (range 1 to 53).

Seven of the 15 authors who responded declined to share study documentation. Reasons given for refusal were as follows: (i) the author had no access to the protocol/SAP and was unable to get in contact with those in possession (n = 2); (ii) author refused due to sponsor policy of not sharing study documents (n = 1); (iii) author refused because they planned to publish further results related to the trial (n = 1); (iv) author refused with no reason given (n = 1); and (v) author stated they would send documents, but did not, despite follow-up emails (n = 2).

Overall, 8 of 96 trials (8%) shared some trial documentation (protocol only [n = 5]; protocol and SAP [n = 1]; excerpt from protocol [n = 1]; research ethics application form [n = 1]). Overall, 6/96 trials (6%) sent a protocol, and 1/96 (1%) sent a SAP. No trials sent both the initial and final version of their SAP. Four authors sent multiple versions of their protocol, though we were unable to verify these included the first and last version. Overall, 4/11 high-impact factor trials (36%) and 4/85 (5%) of PubMed trials shared some trial documentation.

Table 2 Study results

	High impact cohort	PubMed cohort	Total
Publicly available protocol in stage 1			
Yes (excluded from stage 2)	90/101 (89)	15/100 (15)	105/201 (52)
No (included in stage 2)	11/101 (11)	85/100 (85)	96/201 (48)
Replied to email requesting protocol/SAP			
Yes	6/11 (55)	9/85 (11)	15/96 (16)
No	5/11 (45)	76/85 (89)	81/96 (84)
Days until first response (among those responding)			
Median (IQR)	23 (4, 35)	3 (2, 21)	10 (2, 35)
Shared some study documents			
Yes	4/11 (36)	4/85 (5)	8/96 (8)
No	7/11 (64)	81/85 (95)	88/96 (92)
Days until documents shared (among those sharing)			
Median (IQR)	31 (14, 42)	21 (11, 22)	22 (11, 31)
Shared protocol			
Yes	4/11 (36)	2/85 (2)	6/96 (6)
No	7/11 (64)	83/85 (98)	90/96 (94)
Shared statistical analysis plan			
Yes	1/11 (9)	0/85 (0)	1/96 (1)
No	10/11 (91)	85/85 (100)	95/96 (99)
Shared other study documents ^a			
Yes	0/11 (0)	2/85 (2)	2/96 (2)
No	11/11 (100)	83/85 (98)	94/96 (98)
Study protocol available at end of stage 2 (either available publicly in stage 1 or shared upon request in stage 2)			
Yes	94/101 (93)	17/100 (17)	111/201 (55)
No	7/101 (7)	83/100 (83)	90/201 (45)

^a Partial excerpt from a protocol ($n = 1$), research ethics application form ($n = 1$)

In total, we sent out 329 emails during this study (including both our emails asking for study documents, and follow-up emails answering questions from authors). This led to an average rate of 41 emails sent for every one author that shared trial documentation with us.

In stage 1, 52% of trials had a publicly available protocol (89% high-impact cohort; 15% PubMed cohort). After contacting authors in stage 2, this figure increased to 55% (93% high-impact cohort; 17% PubMed cohort), an increase of only 3%.

Undisclosed discrepancies in planned statistical methods

We evaluated 7 of the 8 trials which shared some study documentation. We excluded one trial as there was no detail on the planned statistical methods for the trial's primary outcome in any of the shared documents.

Four of the seven trials (57%) had at least one undisclosed deviation to planned statistical methods (three due to discrepancies in the analysis population and one due to discrepancies in whether covariate adjustment was to be used).

Discussion

Access to protocols and SAPs increases transparency around randomised trials by allowing readers to identify and interpret unplanned changes to study methods which can bias trial results. In stage 1 of this study, we found that only 52% of published trials across two cohorts had a publicly available protocol or SAP. In stage 2, which included 96 trials that did not have a publicly available protocol or SAP, we found that most authors did not share study documentation when requested. Only 6% sent a protocol and 1% sent a SAP. In total, we sent 329 emails as part of this study, meaning that on average, 41 emails were sent for every one trial which shared some form of documentation.

These findings indicate that our approach of contacting authors was an ineffective method of obtaining access to protocols and SAPs. This is concerning, as many trials do not have a publicly available protocol or SAP [16, 18, 21, 24], and so, for a large portion of trials, it is impossible for readers to identify whether authors have followed pre-planned methods, or whether they deviated in order to obtain more favourable results.

Although authors are required to upload information about their trials to trial registry websites such as clinicaltrials.gov, the EU clinical trials register, or the ISRCTN registry, these registries provide very limited information compared to study protocols or SAPs. In particular, most trial registries do not require any information on statistical methods, and so the only place to find this is in the protocol or SAP.

We suggest two simple solutions to improve this situation. First, in addition to information on outcomes, eligibility criteria, and other study methods, clinical trial registries could require authors to upload information about the planned statistical methods. This would ensure that changes to the planned statistical methods could be assessed for all registered trials. This is already a requirement for the Australian New Zealand clinical trials registry, demonstrating the feasibility of this approach. Second, journals could require authors to submit the initial and final version of both their protocol and SAP alongside their trial publication. Some journals have already adopted this policy, which has effectively increased the number of trials with available protocols/SAPs [18]. This policy may partially explain the difference between protocol availability for the high-impact cohort (89%) vs. the PubMed cohort (15%) during stage 1, as some journals in the high-impact cohort had this requirement.

Other researchers have similarly found contacting study authors to be an ineffective method of obtaining trial documents. One recent study found that in cases where the main trial report indicated readers could request the study protocol using a provided email address, protocols were shared for only three of nine requests [25]. Two studies found that sharing of trial data was suboptimal, with data being shared for only 10–46% of requests [26, 27]. Another study found that statistical code was only shared for 15% requests ($n = 7/47$) [28].

This study has some limitations. It involves trials from two different cohorts, one involving high-impact journals and the other involving any journals indexed in PubMed, making overall results more challenging to interpret. We found that the high-impact cohort was more likely to have a publicly available protocol and SAP and also more likely to share these documents when not available. However, the number of trials published in high-impact medical journals is small compared to the overall number of trials published, and so it may be the results of the PubMed cohort (which found only 15% had a publicly available protocol and only 2% shared a protocol when requested) are more generalisable. Second, we estimated that 57% of trials which shared documentation had undisclosed

deviations to the planned analysis approach, which is broadly in line with what others have found [17, 18, 21]. However, this is based on a very small number of trials and so should be interpreted cautiously. Further, it may be that authors who had made changes to the planned methods were less likely to send study documents, so that discrepancies might be more common than what we found here. In addition, we may have had more success receiving SAPs had we emailed the trial statistician directly, instead of the corresponding author (though we note it is not always easy to identify the trial statistician from author lists, or to find their email addresses). Finally, in our email requesting the protocol/SAP, we stated that we wished to evaluate how well authors followed their pre-specified approaches, which may have deterred some authors from sharing.

Conclusion

Most study investigators did not share their unpublished protocols or SAPs upon direct request. Alternative strategies are needed to increase transparency of randomised trials and ensure access to protocols and SAPs.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13063-022-06641-x>.

Additional file 1.

Acknowledgements

Not applicable.

Authors' contributions

DC: wrote first draft of protocol, analysed data, wrote first draft of manuscript. CM: input into protocol and manuscript, contacted corresponding authors, analysed data. SC: input into protocol and manuscript. VP: concept, input into protocol and manuscript. BCK: concept, analysed data, input into protocol and manuscript. The authors read and approved the final manuscript.

Funding

BCK is funded by the UK MRC, grants MC_UU_00004/07 and MC_UU_00004/09. SC is funded by a NIHR advanced fellowship (Reference: NIHR300593).

Availability of data and materials

The datasets used and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Ethics approval and consent to participate

The study protocol was reviewed by the Office of Human Research Ethics at Western University who confirmed that ethical approval was not required.

Consent for publication

Not applicable.

Competing interests

VJ has received consulting fees from AbbVie, Alimentiv Inc (formerly Robarts Clinical Trials), Arena pharmaceuticals, Bristol Myers Squibb, Celltrion, Eli Lilly, Ferring, Fresenius Kabi, GlaxoSmithKline, Genetech, Gilead, Janssen, Merck, Mylan, Pendopharm, Pfizer, Roche, Sandoz, Takeda, Topivert; speaker's fees from, Abbvie, Ferring, Janssen Pfizer, Shire, and Takeda.

Author details

¹Department of Medicine, Division of Gastroenterology, Western University, London, Ontario, Canada. ²Imperial Clinical Trials Unit, Imperial College London, London, UK. ³Department of Epidemiology and Biostatistics, Western University, London, Ontario, Canada. ⁴MRC Clinical Trials Unit at UCL, London, UK.

Received: 21 January 2022 Accepted: 6 August 2022

Published online: 17 August 2022

References

- Chan AW, Altman DG. Identifying outcome reporting bias in randomised trials on PubMed: review of publications and survey of authors. *BMJ*. 2005;330(7494):753.
- Chan AW, Hrobjartsson A, Haahr MT, Gotzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA*. 2004;291(20):2457–65.
- Chan AW, Tetzlaff JM, Altman DG, Laupacis A, Gotzsche PC, Krleza-Jeric K, et al. SPIRIT 2013 statement: defining standard protocol items for clinical trials. *Ann Intern Med*. 2013;158(3):200–7.
- Chan AW, Tetzlaff JM, Gotzsche PC, Altman DG, Mann H, Berlin JA, et al. SPIRIT 2013 explanation and elaboration: guidance for protocols of clinical trials. *BMJ*. 2013;346:e7586.
- Chen T, Li C, Qin R, Wang Y, Yu D, Dodd J, et al. Comparison of clinical trial changes in primary outcome and reported intervention effect size between trial registration and publication. *JAMA Netw Open*. 2019;2(7):e197242-e.
- Dwan K, Altman DG, Arnaiz JA, Bloom J, Chan AW, Cronin E, et al. Systematic review of the empirical evidence of study publication bias and outcome reporting bias. *PLoS One*. 2008;3(8):e3081.
- Goldacre B, Drysdale H, Dale A, Milosevic I, Slade E, Hartley P, et al. COMPare: a prospective cohort study correcting and monitoring 58 misreported trials in real time. *Trials*. 2019;20(1):118.
- Hahn S, Williamson PR, Hutton JL. Investigation of within-study selective reporting in clinical research: follow-up of applications submitted to a local research ethics committee. *J Eval Clin Pract*. 2002;8(3):353–9.
- Li G, Abbade LPF, Nwosu I, Jin Y, Leenus A, Maaz M, et al. A systematic review of comparisons between protocols or registrations and full reports in primary biomedical research. *BMC Med Res Methodol*. 2018;18(1):9.
- Moher D, Hopewell S, Schulz KF, Montori V, Gotzsche PC, Devereaux PJ, et al. CONSORT 2010 explanation and elaboration: updated guidelines for reporting parallel group randomised trials. *J Clin Epidemiol*. 2010;63(8):e1–37.
- Ramagopalan S, Skingsley AP, Handunnetthi L, Klingel M, Magnus D, Pakpoor J, et al. Prevalence of primary outcome changes in clinical trials registered on ClinicalTrials.gov: a cross-sectional study. *F1000Res*. 2014;3:77.
- Rising K, Bacchetti P, Bero L. Reporting bias in drug trials submitted to the Food and Drug Administration: review of publication and presentation. *PLoS Med*. 2008;5(11):e217 discussion e.
- Schulz KF, Altman DG, Moher D, Group C. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ*. 2010;340:c332.
- Vedula SS, Bero L, Scherer RW, Dickersin K. Outcome reporting in industry-sponsored trials of gabapentin for off-label use. *N Engl J Med*. 2009;361(20):1963–71.
- Williamson PR, Gamble C, Altman DG, Hutton JL. Outcome selection bias in meta-analysis. *Stat Methods Med Res*. 2005;14(5):515–24.
- Babu C, Mell L, Lee N, Zakeri K. Public access to protocols of contemporary cancer randomized clinical trials. *Trials*. 2021;22(1):418.
- Chan AW, Hrobjartsson A, Jorgensen KJ, Gotzsche PC, Altman DG. Discrepancies in sample size calculations and data analyses reported in randomised trials: comparison of publications with protocols. *BMJ*. 2008;337:a2299.
- Cro S, Forbes G, Johnson NA, Kahan BC. Evidence of unexplained discrepancies between planned and conducted statistical analyses: a review of randomised trials. *BMC Med*. 2020;18(1):137.
- Dwan K, Altman DG, Clarke M, Gamble C, Higgins JP, Sterne JA, et al. Evidence for the selective reporting of analyses and discrepancies in clinical trials: a systematic review of cohort studies of clinical trials. *PLoS Med*. 2014;11(6):e1001666.
- Greenberg L, Jairath V, Pearse R, Kahan BC. Pre-specification of statistical analysis approaches in published clinical trial protocols was inadequate. *J Clin Epidemiol*. 2018;101:53–60.
- Kahan BC, Ahmad T, Forbes G, Cro S. Public availability and adherence to pre-specified statistical analysis approaches was low in published randomised trials. *J Clin Epidemiol*. 2020;128:29–34.
- Kahan BC, Forbes G, Cro S. How to design a pre-specified statistical analysis approach to limit p-hacking in clinical trials: the pre-SPEC framework. *BMC Med*. 2020;18(1):253.
- Page MJ, McKenzie JE, Forbes A. Many scenarios exist for selective inclusion and reporting of results in randomized trials and systematic reviews. *J Clin Epidemiol*. 2013;66(5):524–37.
- Spence O, Hong K, Onwuchekwa Uba R, Doshi P. Availability of study protocols for randomized trials published in high-impact medical journals: a cross-sectional analysis. *Clin Trials*. 2020;17(1):99–105.
- Sender D, Clark J, Hoffmann TC. Analysis of articles directly related to randomized trials finds poor protocol availability and inconsistent linking of articles. *J Clin Epidemiol*. 2020;124:69–74.
- Naudet F, Sakarovich C, Janiaud P, Cristea I, Fanelli D, Moher D, 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.
- Savage CJ, Vickers AJ. Empirical study of data sharing by authors publishing in PLoS journals. *PLoS One*. 2009;4(9):e7078.
- DeBlanc J, Kay B, Lehrich J, Kamdar N, Valley TS, Ayanian JZ, et al. Availability of statistical code from studies using medicare data in general medical journals. *JAMA Intern Med*. 2020;180(6):905–7.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

