Public availability and adherence to pre-specified statistical analysis approaches was low in published randomised trials

Brennan C Kahan (PhD)[1]*, Tahania Ahmad (MSc)[2], Gordon Forbes (MSc)[3], Suzie Cro (PhD)[4]

[1] MRC Clinical Trials Unit at UCL
[2] Pragmatic Clinical Trials Unit, Queen Mary University of London
[3] Department of Biostatistics and health informatics, Institute of Psychiatry, Psychology & Neuroscience, Kings College London
[4] Imperial Clinical Trials Unit, Imperial College London

* Brennan C Kahan

MRC Clinical Trials Unit at UCL

b.kahan@ucl.ac.uk

Email addresses:

Tahania Ahmad: tahania.ahmad@yahoo.com

Gordon Forbes: gordon.forbes@kcl.ac.uk

Suzie Cro: s.cro@imperial.ac.uk

Abstract

Objective

Pre-specification of statistical methods in clinical trial protocols and Statistical Analysis Plans (SAPs) can help to deter bias from p-hacking, but is only effective if the pre-specified approach is made available.

Study Design and Setting

For 100 randomised trials published in 2018 and indexed in PubMed we evaluated how often a pre-specified statistical analysis approach for the trial's primary outcome was publicly available. For each trial with an available pre-specified analysis, we compared this to the trial publication to identify whether there were unexplained discrepancies.

Results

Only 12 of 100 trials (12%) had a publicly available pre-specified analysis approach for their primary outcome; this document was dated before recruitment began for only two trials. Of the 12 trials with an available pre-specified analysis approach, 11 (92%) had one or more unexplained discrepancies. Only 4/100 trials (4%) stated that the statistician was blinded until the SAP was signed off and only 10/100 (10%) stated the statistician was blinded until the database was locked.

Conclusion

For most published trials, there is insufficient information available to determine whether results may be subject to p-hacking. Where information was available, there were often unexplained discrepancies between the pre-specified and final analysis methods.

Key words: randomised trial, transparency, pre-specification, p-hacking, bias, protocol, statistical analysis plan

Running title: Availability of pre-specified analyses in randomised trials

Word count: **2531**

What is new?

- Few trials made their pre-specified analysis approach publicly available, making it difficult to verify methods were not changed based on trial data (p-hacking)

- Most of the available protocols and Statistical Analysis Plans were dated after recruitment began, so earlier changes could not be ruled out

- Most trials for which it could be assessed had one or more unexplained discrepancy in the statistical methods between the protocol/Statistical Analysis Plan and publication

- Resolution of these issues will require greater emphasis on making protocols and Statistical Analysis Plans publicly available.

Introduction

The statistical methods used to analyse a randomised trial can affect the results; for instance, excluding different participants or using different statistical models can change the size of the estimated treatment effect or p-value (1-14). Selective reporting of results is problematic in randomised trials (4, 6, 7, 13, 15), and, choosing or modifying the planned analysis approach after seeing the trial data to obtain a more favourable result can introduce bias; this is commonly referred to as ”p-hacking” (1-5, 11, 12, 15). Pre-specification of statistical methods is recommended, both to act as a deterrent and help to identify p-hacking (1-5, 11, 12, 15). However, in order for pre-specification to be an effective tool to prevent and identify p-hacking, the pre-specified analysis approach must be publicly accessible; otherwise there is no way for people who are not involved in the trial to verify whether investigators followed their pre-specified approach.

Previous reviews have looked at how often protocols and Statistical Analysis Plans (SAPs) are available for trials published in high-impact general medical journals (15, 16). However, these results are unlikely to be generalizable across all trials; for instance, some high-impact medical journals have policies requiring investigators to include protocols and SAPs as supplementary material alongside the trial publication, which is not the policy at most medical journals. We therefore undertook this study to evaluate, in a random sample of trials published in all journals indexed in PubMed, how often a pre-specified analysis approach was publicly available for the trial’s primary outcome, and how often this approach was modified without explanation.

Methods

We followed the same protocol as a previous review which focussed on trials published in high-impact general medical journals (15), although we used different inclusion/exclusion criteria and used a different search strategy (described below). The protocol is available in the supplementary material.

Search strategy and eligibility criteria

We searched PubMed for randomised trials published in June 2018. Articles were eligible for inclusion if they reported results from a phase 2-4 randomised trial in humans. Exclusion criteria were pilot or feasibility, phase 1, or non-randomised study, secondary analysis of previously published trial, cost-effectiveness as the primary outcome, more than one trial reported in the article, results of an interim analysis, or if the protocol or SAP was not in English. The full search strategy is shown in the supplementary material.

We then randomly selected 100 eligible articles for inclusion in our review. This was achieved by sorting all articles identified from our search into a random order, using a random number generator. One author then screened articles until 100 eligible trials were identified. The list of the 100 included trials is available in the supplementary material.

Data extraction

For each eligible trial, we evaluated whether there was a pre-specified statistical analysis approach for the trial's primary outcome in a publicly available protocol or SAP (i.e. a protocol or SAP that had been published as a peer-reviewed journal article, was included with the trial publication as part of the supplementary material, or was available on the trial's website).

For trials that had an available pre-specified analysis approach, two authors independently assessed whether there were any discrepancies between the pre-specified approach and the trial publication for (i) the analysis population (which participants were included in the analysis, and whether they were analysed according to their allocated treatment arm or not); (ii) the statistical model used for analysis (e.g. a logistic regression model, a mixed-effects linear regression model, or a non-parametric test such as the Wilcoxon test); (iii) any adjustment of baseline covariates; and (iv) how missing data was handled (1, 4).

We evaluated two different types of discrepancies. The first was termed a ‘change’, which meant that investigators had done something differently in the trial publication to what they had pre-specified; for instance, if they pre-specified an intention-to-treat analysis in the trial protocol but then used a per-protocol analysis; or said in the protocol they would adjust for participant age in the analysis but an unadjusted analysis was performed instead.

The second discrepancy was termed an ‘addition’, which occurred if essential details about how the analysis would be implemented were missing from the pre-specified analysis approach. ‘Additions’ were considered problematic as it meant the analysis approach was pre-specified in a manner which still allowed investigators to decide on some details of how the analysis would be implemented after seeing the trial data, which could allow p-hacking (11). Additions occurred if the pre-specified analysis approach (1) contained insufficient information about the proposed analysis; or (2) allowed the investigators to subjectively choose between multiple different potential analyses. Examples of ‘additions’ are if, in the pre-specified analysis approach, investigators: (i) did not specify how they would handle missing data; (ii) pre-specified they would use a per-protocol analysis but did not say which participants they would exclude; or (iii) specified that they would use either a mixed-effects logistic regression model or generalised estimating equations, with the final decision being made after seeing the trial data, but did not specify an objective criteria for deciding between the two approaches.

We classified each discrepancy as either ‘explained’ or ‘unexplained’. We classified discrepancies as being ‘explained’ if they were specified in a subsequent, publicly available version of the protocol or SAP (with or without a justification or rationale for the discrepancy), or if the discrepancy was mentioned in the trial publication. Otherwise, we classified the discrepancy as being ’unexplained’.

We extracted data related to the primary analysis of the primary outcome. We identified a single primary outcome as follows; (a) if one outcome was listed as the primary we used this;

(b) if no outcomes or multiple outcomes were listed as being primary we used the outcome that the sample size calculation was based on; and (c) if no sample size calculation was performed or sample size was calculated for multiple primary outcomes, we used the first clinical outcome listed in the objectives/outcomes section. We identified the primary analysis as follows; (a) if a single analysis strategy was used, or multiple strategies were used with one being identified as primary, we used this; (b) if multiple strategies were used without one being identified as primary, we used the first one presented in the results section.

We extracted data onto a pre-piloted standardised data extraction form (available in the supplementary material). Each trial was evaluated independently by two reviewers, with disagreements being resolved by discussion, or by a third reviewer where disagreement could not be resolved. Where the trial publication referred to supplementary material, a SAP or protocol, the extractor referred to these documents.

Outcomes

Our main outcome measures were (i) the number of trials with a publicly available pre-specified analysis approach for the primary outcome; (ii) the number of trials with no unexplained discrepancies from the publicly available pre-specified analysis approach; and (iii) the total number of analysis elements with an unexplained discrepancy for each trial.

Secondary outcomes were, for each analysis element described earlier (analysis population, statistical model, use of covariates, handling of missing data), (i) the number of trials with at least one unexplained discrepancy (either change or addition); (ii) the number of trials with at least one unexplained change; and (iii) the number of trials with at least one unexplained addition.

Results

Search results and characteristics of included studies

Our search identified 1382 results (figure 1). We then identified 100 eligible trials after reviewing 327 randomly ordered articles. Trial characteristics are shown in table 1.

Protocols were available for 15 trials (15%) (8 published, 7 as supplementary material with publication, 1 on a website; one trial had two protocols available from different sources). SAPs were available for 3 trials (3%) (0 published, 2 as supplementary material with publication, 1 on a website). The three trials with a SAP also had an available protocol. Of the 15 trials with an available protocol, the earliest version available was dated before the start of recruitment for 3 (20%) trials, 7 (47%) were dated after recruitment had begun, and for 5 (33%) trials this was unclear (three trials had a protocol with no date and two trials did not report the recruitment start date). Of the three trials with an available SAP, the earliest version of the SAP was dated after recruitment began for all three (100%) trials (for two trials it was dated over 1 year after recruitment had ended).

Of the 100 trials included in this review, only 4 (4%) stated in the trial publication, protocol, or SAP that the statistician was blinded until the SAP was signed off and only 10 (10%) stated the statistician was blinded until the database was locked.

Availability of pre-specified analysis approach

Overall, 12 of 100 trials (12%) had a publicly available pre-specified analysis approach for the primary outcome (table 2). 85 trials did not have an available protocol or SAP, and 3 trials had a protocol with no information on the analysis and no publically available SAP. The document containing the pre-specified analysis approach was dated before the start of recruitment for 2 of 12 (17%) trials, during recruitment in 1 (8%) trial (7 months post-

recruitment beginning), and after the end of recruitment in 5 (42%) trials (median 6 months post-recruitment completion, IQR 3 to 8). In 4 trials (33%) it was unclear when the document was dated in relation to the start of the trial (for two trials no date was available for the protocol; for two trials no date was available for the commencement of recruitment).

Comparison of pre-specified and conducted statistical analysis approach

Of the 12 trials with an available pre-specified analysis approach, only 1 (8%) did not have any unexplained discrepancies (it had discrepancies, but all were explained) (table 2). A further 11 trials (92%) had one or more unexplained discrepancies; the total number of discrepancies across the 11 trials was n=23 (table 3). For 10 of these discrepancies, the investigators made a change to their pre-specified methods, for 7 discrepancies the investigators did not provide any information in the protocol for that analysis element, and for 6 discrepancies the investigators pre-specified their methods in a manner which could have allowed them to choose the analysis method after seeing the trial dataset in order to give a more favourable result.

Unexplained discrepancies were most common for the use of covariates (n=8/12, 67%) and the analysis model (n=6/12, 50%), though were also high for the handling of missing data (n=5/12, 42%) and the analysis population (n=4/12, 33%). The total number of changes and additions for each analysis element is available in table 4.

Post-hoc analysis of trial registry entries

We performed a post-hoc analysis to evaluate whether any trial registry entries contained prospective information on the planned statistical analysis approach for the trial's primary outcome. We found that 27 trials did not have an evaluable trial registry entry (no registration number listed [n=20]; listed registration number could not be found on the relevant registry

website [n=4]; trial registry not in English [n=3]). None of the 73 evaluable trial registry entries (0/73, 0%) contained any prospective information on the planned analysis approach for the trial's primary outcome. Three registry entries contained retrospective information, added after trial was finished, describing how data had been analysed.

Discussion

In this review of trials indexed in PubMed, we found that very few (12%) had a publicly available pre-specified analysis approach. Furthermore, because most documents containing the pre-specified analysis approach were dated after the trial had begun (or, in some cases, after the trial was completed), it was often difficult to ascertain whether the analysis approach in these documents had already been modified from the pre-trial protocol.

Of the trials that did have a pre-specified analysis approach available, we found that almost all of them (92%) had unexplained discrepancies. These typically involved making undisclosed changes to the pre-planned methods, or pre-specifying the methods in a manner which allowed p-hacking (11). Because of poor reporting around blinding of statisticians and access to data, it was often difficult to ascertain whether changes were made before or after access to the trial data.

Previous reviews evaluating availability of protocols and SAPs in general medical journals have found much higher rates of availability than we did (16). Spence *et al* reviewed 364 trials from five high-impact medical journals and found that protocols and SAPs were available in 82% and 50% of trials respectively (16). In a separate review of 101 trials from six high-impact medical journals, we found that 88% had a publicly available pre-specified analysis approach for the primary outcome (15). The reasons for these differences are likely due in part to differences in policy between some high-impact medical journals vs. other journals (e.g. the requirement to publish the protocol and SAP alongside the trial). However,

the rate of unexplained discrepancies we found in this review was broadly in line with previous reviews (2, 6, 7, 15).

Resolution of these issues will require greater emphasis on making protocols and Statistical Analysis Plans publicly available, and better reporting around blinding and data access of statisticians and investigators. Journal editors could require that the first and last version of the protocol and Statistical Analysis Plan be submitted alongside the trial report and be published as supplementary material; this is the approach taken by the *New England Journal of Medicine*, and it has been shown to be very effective (15). Additionally, trial registries could require authors to disclose the planned analysis approach for their primary outcome at the time of registration. This would ensure that most trials could be assessed for inappropriate changes to the analysis approach. Next, authors could ensure they do not pre-specify their analysis approach in a way which allows p-hacking; the Pre-SPEC framework, which has recently become available, provides guidance on this exact issue (11). Finally, editors could require that authors report any changes to the pre-specified methods, as well as disclose the blinding status of investigators at the time of each change (authors could also do so voluntarily).

This review had some limitations. We only identified 12 articles which had a protocol containing a pre-specified analysis approach, and so our estimates for the rates of unexplained discrepancies were based on a small number of trials. However, given that previous reviews have identified similar rates of discrepancies (2, 6, 7, 15), we are confident that the true rate of unexplained discrepancies is large enough to warrant concern. Second, we have assumed that the protocols and SAPs made available alongside the published trial results do represent valid pre-specified analysis approaches. However, this relies on the assumption that the dates of such documents are accurate, i.e. they were indeed written and signed off before analysis began. This assumption is unverifiable, and so it is possible that some of these documents were not in fact written prior to the commencement of final

analysis. Third, although we found that most trials did have discrepancies, we were not able to ascertain whether these discrepancies were due to p-hacking, or whether they occurred for other reasons. Finally, we included only studies published in English, and so results may not be generalisable to other settings.

Conclusion

For most published trials there is insufficient information available to determine whether results may be subject to bias due to p-hacking. Where information was available, there were often unexplained discrepancies between the pre-specified and final analysis methods.

Figure 1 – Flow diagram of article selection

Table 1 – Characteristics of included trials (data are n [%])

Characteristic	Total (N=100)
Funding (n, %)	
Pharmaceutical	9 (9%)
Other for profit company	3 (3%)
Government	20 (20%)
Charity sector	5 (5%)
Other	24 (24%)
Multiple funding sources	8 (8%)
Unclear funding source	31 (31%)
Type of intervention (n, %)	
Pharmacologic	48 (48%)
Surgical	10 (10%)
Psychosocial/behavioural/educational	12 (12%)
Other	29 (29%)
Multiple intervention types	1 (1%)
Cluster trial (n, %)	2 (2%)
Factorial trial (n, %)	1 (1%)
Crossover trial (n, %)	3 (3%)
Non-inferiority (n, %)	5 (5%)
No. of treatment arms (n, %)	
Two	81 (81%)
Three or more	19 (19%)
Sample size (median, IQR)	99.5 (60, 174)

IQR=interquartile range

Table 2 - Unexplained discrepancies in statistical methods between the pre-specified approach and the trial publication (data are n [%])

Measure	Total (N=12)
Unexplained discrepancies	
None[a]	1 (8%)
≥1	11 (92%)
Number of unexplained discrepancies[b]	
0[a]	1 (8%)
1	3 (25%)
2	4 (33%)
3	4 (33%)
Analysis elements with unexplained discrepancies	
Analysis population	4 (33%)
Analysis model	6 (50%)
Use of covariates	8 (67%)
Handling of missing data	5 (42%)

[a]no discrepancies (n=0), explained discrepancies only (n=1).

[b]Calculated as number of analysis elements with an unexplained discrepancy (analysis elements are: (i) analysis population; (ii) analysis model; (iii) use of covariates; and (iv) handling of missing data).

Table 3 - Description of unexplained discrepancies (n=12) (data are n [%])

Measure	Total (N=12)
Analysis population	
No information given in original analysis approach[a]	1 (8%)
Changed from pre-specified population by specifying additional exclusions[b]	3 (25%)
Analysis model	
No information given in original analysis approach[a]	2 (17%)
Original analysis approach allowed analyst to subjectively choose final model based on trial dataset[a]	1 (8%)
Changed from pre-specified model[b]	3 (25%)
Use of covariates	
No information given in original analysis approach[a]	1 (8%)
Original analysis approach allowed analyst to subjectively choose covariates based on trial dataset[a]	4 (33%)
Changed from pre-specified approach[b]*	3 (25%)
Handling of missing data	
No information given in original analysis approach[a]	3 (25%)
Analysis approach allowed analyst to subjectively choose final multiple imputation model based on trial dataset[a]	1 (8%)
Changed from imputation to complete case[b]	1 (9%)

[a] Classified as an 'addition'

[b] Classified as a 'change'

*N=2 changed from unadjusted (pre-specified) to adjusted analysis; N=1 changed from adjusted (pre-specified) to unadjusted analysis

Table 4 – Unexplained changes and additions (N=12)

Characteristic	N (%)
Analysis population	
Unexplained discrepancy (change or addition)	4 (33%)
Unexplained addition	1 (8%)
Unexplained change	3 (25 %)
Analysis model	
Unexplained discrepancy (change or addition)	6 (50%)
Unexplained addition[c]	3 (25%)
Unexplained change[c]	3 (25%)
Covariates	
Unexplained discrepancy (change or addition)	8 (67%)
Unexplained addition	5 (42%)
Unexplained change	3 (2%)
Missing data	
Unexplained discrepancy (change or addition)	5 (42%)
Unexplained addition	4 (33%)
Unexplained change	1 (8%)

References

1. 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.
2. 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.
3. Gamble C, Krishan A, Stocken D, Lewis S, Juszczak E, Dore C, et al. Guidelines for the Content of Statistical Analysis Plans in Clinical Trials. Jama-Journal of the American Medical Association. 2017;318(23):2337-43.
4. 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.
5. ICH Harmonised Tripartite Guideline. Statistical principles for clinical trials. International Conference on Harmonisation E9 Expert Working Group. Stat Med. 1999;18(15):1905-42.
6. 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.
7. Dwan K, Altman DG, Cresswell L, Blundell M, Gamble CL, Williamson PR. Comparison of protocols and registry entries to published reports for randomised controlled trials. Cochrane Database Syst Rev. 2011(1):MR000031.
8. Porta N, Bonet C, Cobo E. Discordance between reported intention-to-treat and per protocol analyses. J Clin Epidemiol. 2007;60(7):663-9.
9. Prakash A, Risser RC, Mallinckrodt CH. The impact of analytic method on interpretation of outcomes in longitudinal clinical trials. Int J Clin Pract. 2008;62(8):1147-58.
10. Saquib N, Saquib J, Ioannidis JP. Practices and impact of primary outcome adjustment in randomized controlled trials: meta-epidemiologic study. BMJ. 2013;347:f4313.
11. 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. https://arxivorg/abs/190704078. 2020.
12. 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.
13. 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.
14. Sterne JAC, Savovic J, Page MJ, Elbers RG, Blencowe NS, Boutron I, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. BMJ. 2019;366:l4898.
15. 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.
16. 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. 2019:1740774519868310.

Funding statement
None.

Author contributions

BCK: conceptualisation, methodology, investigation, writing – original draft, supervision, project administration. TA: Writing - Review & Editing, investigation. GF: conceptualisation, methodology, investigation, Writing - Review & Editing. SC: conceptualisation, methodology, formal analysis, investigation, data curation, Writing - Review & Editing.

Declaration of interest

Declarations of interest: none