BMJ 2016;355:i5543 doi: 10.1136/bmj.i5543

FEATURE

REPORTING OUTCOMES

Is this trial misreported? Truth seeking in the burgeoning age of trial transparency

One group's efforts to monitor misreporting of outcomes has irritated several medical journals, which argue that the differences discovered are not clinically important. So how seriously should we be taking it? **Peter Doshi reports**

Peter Doshi *associate editor, The BMJ*

Glimpses inside SmithKline Beecham's secret clinical trials programme for the antidepressant paroxetine began in the early 2000s. Amid a growing storm over the safety of selective serotonin reuptake inhibitors for children, a leaked memo revealed by the BBC's *Panorama* programme[1] depicted a company trying to manage the unfavourable results of two important trials. "It would be commercially unacceptable to include a statement that efficacy had not been demonstrated, as this would undermine the profile of paroxetine," the memo read. A reviewer for the US Food and Drug Administration considered both trials as "failed."[2]

But then there was the public face of the data. One of the two trials–Study 329–was published in the peer reviewed literature.[3] The manufacturer told its sales representatives that the "landmark study … demonstrates REMARKABLE efficacy and safety."[4]

Study 329 has become a classic example of what is known as outcome reporting bias, in which trial authors selectively present trial results leading, almost inevitably, to a rosier picture than would have occurred had the trial been reported according to the original protocol. The data showed no difference between paroxetine and placebo for all eight of the originally specified outcomes of interest, and an increase in harms.[5-7] Yet the 2001 trial publication[3] reported on four outcomes not specified in the protocol (all of which had statistically significant differences) and concluded that paroxetine was "generally well tolerated and effective." Put simply, the goalposts set when the trial commenced had moved by the time the trial was reported.

Some may have suspected that Study 329 was an anomaly. But in May 2004, an empirical analysis comparing the protocols of randomised trials with their associated publications suggested otherwise.[8] Nearly two thirds of protocols did not match the publication in what seems like the most fundamental way: the study's primary outcomes. The protocol set out what the investigators hoped to study, but in 40 out of 82 trials the specified primary outcomes were not presented as such in the final publication. Sometimes they were reported as non-primary. Other times they were simply missing from the publication. In 11 further trials, outcomes not mentioned in the protocol suddenly appeared as the "primary outcome" in the publication.

Outraged editors

The long awaited goal of universal registration of trials now seemed achievable,[9] and medical journal editors issued an ultimatum: preregister your trial or forgo publication in our pages.[10] "Honest reporting begins with revealing the existence of all clinical studies, even those that reflect unfavourably on a research sponsor's product," a group of influential editors declared. "Unfortunately, selective reporting of trials does occur, and it distorts the body of evidence available for clinical decision-making." The declaration had enormous impact,[11] and public trial registers remain a key mechanism to prevent investigators from hiding or spinning unfavourable results.[12]

But more than a decade on, a small project from Oxford University's Centre for Evidence Based Medicine seems to have journal editors eating their own words, with some of the world's most powerful editors arguing that strict adherence to the registry entry or trial protocol may not always make sense.

Enter COMPare

Launched by Ben Goldacre and colleagues in late 2015, the COMPare[13] project injected an old idea with an activist twist. Rather than simply gauging the prevalence of selective reporting of outcomes, they would write a letter to the editor for any trial containing "switched outcomes," to see how journals would react. Would journals welcome COMPare's detective work, promptly issuing corrections, or would editors resist?

pdoshi@bmj.com

Data supplements on bmj.com (see http://www.bmj.com/content/355/bmj.i5543?tab=related#datasupp)
NEJM response to COMPare

Over six weeks, Goldacre's team analysed 67 trials published across five leading medical journals–*Annals of Internal Medicine*, *The BMJ*, *JAMA*, *Lancet*, and *New England Journal of Medicine (NEJM)*.[14] Only nine trials, according to COMPare, were "perfect," meaning the publication reported all prespecified outcomes and flagged for readers any changes from the protocol (such as introducing outcomes after the trial started). Between them the 58 offending trials failed to report 354 specified outcomes and "silently added" an almost equal number (357). The team dispatched 58 letters to the journal editors seeking correction to the published record. However, few corrections resulted.

Journal dissent

"Upon receipt of COMPARE's initial communication, our editorial team (comprised of physicians and statisticians) thoroughly re-reviewed materials associated with the articles," *Annals'* editor in chief, Christine Laine, told *The BMJ*. "We concluded that the information reported in the articles audited by COMPare accurately represents the scientific and clinical intent detailed in the protocols." In notices posted to the journal's website, *Annals* editors acknowledged the good intentions of COMPare but warned people to be wary.

"Until the COMPare Project's methodology is modified to provide a more accurate, complete and nuanced evaluation of published trial reports, we caution readers and the research community against considering COMPare's assessments as an accurate reflection of the quality of the conduct or reporting of clinical trials."[15] In an interview, Laine pointed to a trial her journal and the COMPare team had tussled over for months (box 1).

NEJM also flunked COMPare's test. The journal's editors refused to publicly engage with the group, rejecting all of COMPare's 20 letters, most with a tightly worded statement saying "we have not identified any clinically meaningful discrepancies between the protocol and the published paper that are of sufficient concern to require correction of the record."[20] But in one case, *NEJM* offered a detailed defence (box 2).

NEJM's response has a certain irony. The journal's editor, Jeffrey Drazen, has been a longtime supporter of trial registration,[10-26] and *NEJM* was the first of its peers to publish trial protocols alongside trial publications, a practice now followed by *JAMA*, *The BMJ*, and most recently–in part thanks to COMPare[27]—*Annals*.

The BMJ did not escape criticism but ultimately got a green light. COMPare sent rapid responses for two of the three trials evaluated,[28 29] one of which led to a correction.[30] It was "an example of best practice," the group said in a blog.[31]

What about *JAMA* and the *Lancet*? *JAMA* rejected all 11 letters the group sent,[32] and the *Lancet* rejected some but published others.[33] COMPare has published the original letters and its correspondence with editors on its website.

Lost in a maze of detail

The disputes are so detail oriented that my eyes crossed trying to follow what at times feels like a frenzied match of ping pong, each side's latest rejoinder seeming to rebut their opponents' last counterpoint. In one trial, COMPare's datasheet shows that the group counted 83 secondary outcomes.[34] Yet in its letter[35] to *NEJM*, it mentions "29 pre-specified secondary outcomes" while the last entry before patient enrolment on ClinicalTrials.gov lists just two.[36] By my count—based on the protocol[37] dated before patient enrolment and posted on NEJM.org—there were over 200.

I shared the case with Curtis Meinert, Johns Hopkins professor and former editor of the journal *Clinical Trials*. Meinert guessed that fear of future accusations of bias was driving trialists to bad practices. "They don't want to be accused of looking at something they didn't specify, so they specify every freaking thing."

What is the clinical significance?

I asked COMPare to elaborate on clinical impact. Beyond the technical violations of best practices in trial reporting, did the team discover any misreporting that concerned them from a clinical perspective? For COMPare, this question was out of scope.

"On clinical impact, we've deliberately not examined that. We set out to assess whether outcomes are correctly reported in journals listed as endorsing CONSORT," the team told me, referring to the well established guidelines for proper reporting of randomised trials. "We deliberately steered away from any value judgments on why someone reported something other than their prespecified outcomes, because we needed our letters to be factual, unambiguous, and all comparable between trialists and journals."

But clinical and value judgments were central to journal editors' defence.

"COMPare objects to the format in which the data are communicated, but COMPare is silent about whether they dispute the key clinical message of the article," *NEJM* editor Jeffrey Drazen said through a spokesperson.

Annals echoed this sentiment: "COMPare's methodology suggests a lack of recognition of the importance of clinical judgment when assessing pre-stated outcomes."

Adhering to the original plan

The challenge of establishing outcome "switching" begins with determining trialists' prespecified outcomes. But which document—protocol, statistical analysis plan, registry entry, or some combination of the above—details trialists' true intentions?

COMPare's methods prioritise[38] protocols over registry entries, a practice that troubles Elizabeth Loder, head of research at *The BMJ*, which requires reporting according to the registry entry. "I see a worrying trend away from trial registries to protocols … People seem to think that as long as you've registered the trial and given notice of its existence, the details don't matter so much."

Annals said that sometimes editors are faced with "a choice between an incomplete or non-updated registry record, with a reliable date stamp, and a more detailed and updated protocol document."

Such situations deeply trouble Deborah Zarin, a long time advocate of trial registration and director of ClinicalTrials.gov, who believes that trial registration is the foundation of the trial reporting system.[39] (Zarin is on the advisory board of OpenTrials, another project directed by Goldacre.)

Even with trialists' last testament established, complexity remains. Having herself conducted similar analyses that compare outcomes reported across different sources,[40 41] Zarin noted[42] that a key challenge is disagreement in the community over how detailed outcome prespecification must be.

Box 1: COMPare versus *Annals of Internal Medicine*

COMPare says it found evidence of incorrectly reported outcomes[16 17] in a trial of the Alexander technique and acupuncture for chronic neck pain[15]:

"There were 3 pre-specified primary outcomes (one score at 3 timepoints). Although data for all 3 time points is reported in the paper, it is also incorrectly stated that the primary endpoint was at 12 months, when no such ranking is given in the registry entry," COMPare wrote to *Annals* last November.

Annals does not dispute that the registry entry fails to denote one of the three timepoints as "primary." But this is no showstopper for the editors, who accept the validity of the published trial protocol (which specifies the 12 month timepoint as primary), even though it was published over a year after the trial started.[18] *Annals* told *The BMJ*: "It is also important to note that most patients and clinicians would consider the longer-term outcome (12 months) to be the more clinically important" than the shorter term outcomes.

But COMPare contends the published protocol is irrelevant and "cannot, by definition, be a source of pre-specified outcomes."[19]

COMPare also alleges that the trial published by *Annals* omits 11 of 17 prespecified measures of adverse events. *Annals* rejects the charge, noting that the article included a summary of serious and non-serious adverse events experienced during the full trial.

Box 2: COMPare versus *NEJM*

COMPare and *NEJM* traded lengthy rejoinders[21 22] over a trial of a hepatitis C treatment (see web extra for most recent *NEJM* response).[23] COMPare initially[24] alleged that a primary outcome was relabelled as secondary in the journal article and that several protocol specified secondary outcomes went unreported, among other issues.

The first point related to the reporting of adverse events. COMPare contends that this outcome was "incorrectly reported" as it was initially defined as one of two primary outcomes but then labelled a secondary outcome in the publication. *NEJM* does not dispute the facts but notes that one primary endpoint was for an efficacy variable and the other for safety. Although labelled "secondary" in the Methods section of the publication, the "safety results are prominently reported in the published paper," *NEJM* says. They appear in the first sentence of the paper's Safety section under Results.

A second point in dispute is the reporting of an efficacy endpoint declared in the protocol that the COMPare team says was not reported in the publication. But *NEJM* counters that "information relevant to" the outcome (specified as a secondary outcome in the registry) was reported in the paper, although it was not labelled a secondary endpoint. Readers wanting to determine the whether or not the outcome was prespecified as primary or secondary would have to consult the protocol or registry entry, both publicly available.

Then there is a secondary outcome that both COMPare and *NEJM* agree was not reported but disagree on whether this was acceptable. According to *NEJM*, the outcome could not have been reported because at the time of the trial's publication, adjudication of the data, which involved data from 24 weeks after treatment, was still underway. COMPare says that this fact should have been disclosed in the manuscript.

COMPare told *The BMJ* that the *NEJM*'s response missed the point. "It only attempts to move the goalposts and discuss other issues, such as whether sofosbuvir is effective."

"From my perspective, prespecification sets the foundation for a statistical analysis; if there is no firmly prespecified outcome measure, then it's unclear to me what any reported P value means," she said, noting that ClinicalTrials.gov now requires investigators to delineate "the specific metric and the time point(s)" for all outcome measures.

Yet some trialists do not sufficiently define outcomes until after the data are collected (but not unmasked) and they finalise the statistical analysis plan, a situation the FDA apparently accepts.

The FDA told *The BMJ* that it "does not require the statistical analysis plan to be in place prior to initiation of the trial."

Annals stated that "prespecification is certainly important, but holding blindly to this prespecification may not be good science. Sometimes it becomes clear to informed clinicians and scientists who understand the area that what was prespecified is not going to provide patients and clinicians with the information they need."

CONSORT guidelines author David Moher agrees that prespecification should not necessarily be held "to the altar,"[19] but said, "What's essential and missing is the notion of transparency. The notion that we've done this and we've changed something … We really need to know if somebody has changed something." He commended COMPare for conducting its audit and providing feedback, some of which "is stressful to journals."

Meinert, however, was more sympathetic to investigators. "Having spent my life in the trenches, I know things never go the way you planned." There are all sorts of sound reasons for investigators changing what they are interested in studying as a trial progresses, he said, and he was "wary of all the reporting bias types of arguments." We're too quick to assume that something is biased, he said. "Most of the stuff you're looking at is probably carelessness rather than bias."

"We do not regard outcome switching as evidence of misconduct," COMPare wrote in a blog. "Our impression is that this is a structural and cultural failing to take outcome switching seriously … However, we do believe that permissiveness around outcome switching, where it is done thoughtlessly, will give cover to those who do engage in outcome switching to deliberately exaggerate their results."[43]

Scope of the problem

Whether cases like Study 329 are exceptional events or the tip of the iceberg may ultimately matter less than the harm even single cases can bring, from the prescribing of therapies thought to be proved safe and effective that are in fact unsafe and ineffective, to the loss in trust in our systems for establishing reliable medical evidence.

"Things are not quite as black and white as COMPare seems to think," Loder commented. And detecting misreporting of serious consequence may take a far deeper investigation of specific cases. But ultimately, the effort is critical. "Watchdog groups like COMPare play an important role in educating the public and motivating journals and authors to do better," Loder said.

I thank Tom Jefferson for helpful comments.

Competing interests: I have read and understood BMJ policy on declaration of interests and declare I am a colleague of many people in this article and those involved in the COMPare project. *The BMJ* is a cofounder of the AllTrials campaign, together with Ben Goldacre, and a BMJ editor is on the advisory board of the OpenTrials project that Goldacre directs.

Provenance and peer review: Commissioned; externally peer reviewed.

1 GSK knew Seroxat wasn't 'effective' on children. *BBC News* 2004 Feb 3.http://news.bbc.co.uk/2/hi/programmes/panorama/3454873.stm

2 FDA. Clinical review: paroxetine. 2002. www.fda.gov/downloads/Drugs/DevelopmentApprovalProcess/DevelopmentResources/UCM371260.pdf

3 Keller MB, Ryan ND, Strober M, et al. Efficacy of paroxetine in the treatment of adolescent major depression: a randomized, controlled trial. *J Am Acad Child Adolesc Psychiatry* 2001;40:762-72. doi:10.1097/00004583-200107000-00010pmid:11437014.
4 Memo to sales representatives. 2001. http://www.healthyskepticism.org/documents/documents/20010816Hawkinstoreps.pdf
5 GSK. Paroxetine—paediatric and adolescent patients.http://www.gsk.com/en-gb/media/resource-centre/paroxetine/paroxetine-paediatric-and-adolescent-patients/
6 Jureidini J, McHenry L, Mansfield P.Clinical trials and drug promotion: selective reporting of study 329. *Int J Risk Saf Med* 2008;20:73-81.
7 Le Noury J, Nardo JM, Healy D, et al. Restoring Study 329: efficacy and harms of paroxetine and imipramine in treatment of major depression in adolescence. *BMJ* 2015;351:h4320. doi:10.1136/bmj.h4320pmid:26376805.
8 Chan A-W, Hróbjartsson A, Haahr MT, Gøtzsche PC, Altman DG. Empirical evidence for selective reporting of outcomes in randomized trials: comparison of protocols to published articles. *JAMA* 2004;291:2457-65. doi:10.1001/jama.291.20.2457. pmid:15161896.
9 Rennie D. Trial registration: a great idea switches from ignored to irresistible. *JAMA* 2004;292:1359-62. doi:10.1001/jama.292.11.1359pmid:15355937.
10 DeAngelis CD, Drazen JM, Frizelle FA, et al. International Committee of Medical Journal Editors. Clinical trial registration: a statement from the International Committee of Medical Journal Editors. *JAMA* 2004;292:1363-4. doi:10.1001/jama.292.11.1363. pmid:15355936.
11 Zarin DA, Tse T, Sheehan J. The proposed rule for U.S. clinical trial registration and results submission. *N Engl J Med* 2015;372:174-80. doi:10.1056/NEJMsr1414226. pmid:25539444.
12 Dickersin K, Rennie D. The evolution of trial registries and their use to assess the clinical trial enterprise. *JAMA* 2012;307:1861-4. doi:10.1001/jama.2012.4230. pmid:22550202.
13 COMPare. Tracking switched outcomes in clinical trials. 2016. www.COMPare-trials.org
14 COMPare. Post-hoc “pre-specification” and undeclared deferral of results: a broken record in the making. 29 Jan 2016.http://compare-trials.org/blog/post-hoc-pre-specification-and-undeclared-separation-of-results-a-broken-record-in-the-making/
15 MacPherson H, Tilbrook H, Richmond S, et al. Alexander technique lessons or acupuncture sessions for persons with chronic neck pain: a randomized trial. *Ann Intern Med* 2015;163:653-62. doi:10.7326/M15-0667pmid:26524571.
16 COMPare. Trial 25 assessment. https://docs.google.com/spreadsheets/d/1qVmaNwxclrTtZzUh2YaecFj7bOIAIndvRSoRkHyQMBQ/edit?pref=2&pli=1#gid=0 &
17 COMPare. Trial 25 letter. https://docs.google.com/document/d/1M_KNDBeSh4WVvWyz4nobpon4-LYGNNOIG-dXChChUTo/edit?pref=2&pli=1
18 MacPherson H, Tilbrook HE, Richmond SJ, et al. Alexander technique lessons, acupuncture sessions or usual care for patients with chronic neck pain (ATLAS): study protocol for a randomised controlled trial. *Trials* 2013;14:209. doi:10.1186/1745-6215-14-209pmid:23841901.
19 Slade E, Dryden H, Goldacare B. Electronic response to: MacPherson et al Ann Intern Med 2015 http://annals.org/article.aspx?articleid=2467961.
20 COMPare. How did NEJM respond when we tried to correct 20 misreported trials? 25 Feb 2016. http://compare-trials.org/blog/how-did-nejm-respond-when-we-tried-to-correct-20-misreported-trials/
21 NEJM replies. https://www.propublica.org/documents/item/2783316-NEJM-replies.html
22 COMPare reply. https://www.propublica.org/documents/item/2783318-Compare-Reply.html
23 Feld JJ, Jacobson IM, Hézode C, et al. ASTRAL-1 Investigators. Sofosbuvir and velpatasvir for HCV genotype 1, 2, 4, 5, and 6 infection. *N Engl J Med* 2015;373:2599-607. doi:10.1056/NEJMoa1512610pmid:26571066.
24 COMPare. Trial 49 letter. https://docs.google.com/document/d/1YvUqNW7sXhaWFTd4iQhrD6dVbfc3NCHZcKVowbXwZUM/edit
25 Drazen JM, Zarin DA. Salvation by registration. *N Engl J Med* 2007;356:184-5. doi:10.1056/NEJMe068291pmid:17215537.
26 DeAngelis CD, Drazen JM, Frizelle FA, et al. International Committee of Medical Journal Editors. Is this clinical trial fully registered? A statement from the International Committee of Medical Journal Editors. *JAMA* 2005;293:2927-9. doi:10.1001/jama.293.23.jed50037pmid:15911838.
27 McCook A. In major shift, medical journal to publish protocols along with clinical trials. Retraction Watch 13 May 2016. http://retractionwatch.com/2016/05/13/in-major-shift-annals-to-publish-protocols-along-with-clinical-trials/
28 Drysdale HM. Discrepancies between pre-specified and reported outcomes in Gilbody et al. Electronic response to Gilbody et al Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial BMJ 2015. http://www.bmj.com/content/351/bmj.h5627/rr-12 &
29 Drysdale HM, Goldacre B. Response to the authors: Outcomes consist of time-points and methods of measurement, as stated in CONSORT. Electronic response to van der Aa et al Stepped care for depression and anxiety in visually impaired older adults: multicentre randomised controlled trial. *BMJ* 2015.http://www.bmj.com/content/351/bmj.h6127/rr-1
30 Computerised cognitive behaviour therapy (cCBT) as treatment for depression in primary care (REEACT trial): large scale pragmatic randomised controlled trial. *BMJ* 2016;352:i195. http://www.bmj.com/content/352/bmj.i195.pmid:26759375.
31 Drysdale H. REEACT in the BMJ: Good science in action. 20 Jan 2016. http://compare-trials.org/blog/reeact-in-the-bmj-good-science-in-action/
32 Goldacre B. JAMA reject correction letters on all trials they have misreported24 Mar 2016.http://compare-trials.org/blog/jama-reject-all-correction-letters/
33 Goldacre B. The Lancet and COMPare: why journals should address outcome switching themselves. 29 Apr 2016. http://compare-trials.org/blog/the-lancet-and-compare-why-journals-should-address-outcome-switching-themselves/
34 COMPare. Trial 039 assessment. https://docs.google.com/spreadsheets/d/1CBHSYsvqK5C1YztzA9L6i0o5KkbRG2zjoTpHIMz87uA/edit#gid=0
35 COMPare. Letter to *NEJM* 2 Dec 2015. https://docs.google.com/document/d/1eDeSFToRbaaz7y1WYFNUpHkpRO6rZPh6aiCXaCXZ_Ys/edit
36 View of NCT01751906 on 2013_01_08. https://clinicaltrials.gov/archive/NCT01751906/2013_01_08
37 Protocol for: Ellis SG, Kereiakes DJ, Metzger DC, et al. Everolimus-eluting bioresorbable scaffolds for coronary artery disease. *N Engl J Med* 2015.http://www.nejm.org/doi/suppl/10.1056/NEJMoa1509038/suppl_file/nejmoa1509038_protocol.pdf
38 Centre for Evidence Based Medicine Outcome Monitoring Project protocol. (last amended 18 May 2016). http://compare-trials.org/wp-content/uploads/2016/01/COMPare-Protocol-18.5.2016.pdf
39 Zarin DA, Tse T. Sharing individual participant data (IPD) within the context of the trial reporting system (TRS). *PLoS Med* 2016;13:e1001946. doi:10.1371/journal.pmed.1001946. pmid:26784335.
40 Hartung DM, Zarin DA, Guise JM, McDonagh M, Paynter R, Helfand M. Reporting discrepancies between the ClinicalTrials.gov results database and peer-reviewed publications. *Ann Intern Med* 2014;160:477-83. doi:10.7326/M13-0480pmid:24687070.
41 Schwartz LM, Woloshin S, Zheng E, Tse T, Zarin DA. ClinicalTrials.gov and Drugs@FDA: A comparison of results reporting for new drug approval trials. *Ann Intern Med* 2016;165:421-30. doi:10.7326/M15-2658pmid:27294570.
42 Zarin DA, Tse T. Trust but verify: trial registration and determining fidelity to the protocol. *Ann Intern Med* 2013;159:65-7. doi:10.7326/0003-4819-159-1-201307020-00011pmid:23817705.
43 Goldacre B. Where does Annals of Internal Medicine stand on outcome switching? A detailed response. 20 Jan 2016. http://compare-trials.org/blog/where-does-annals-of-internal-medicine-stand-on-outcome-switching-a-detailed-response/#more-232