

Comment

Check for publication integrity before misconduct

Andrew Grey, Mark J. Bolland, Alison Avenell, Andrew A. Klein & C. K. Gunsalus

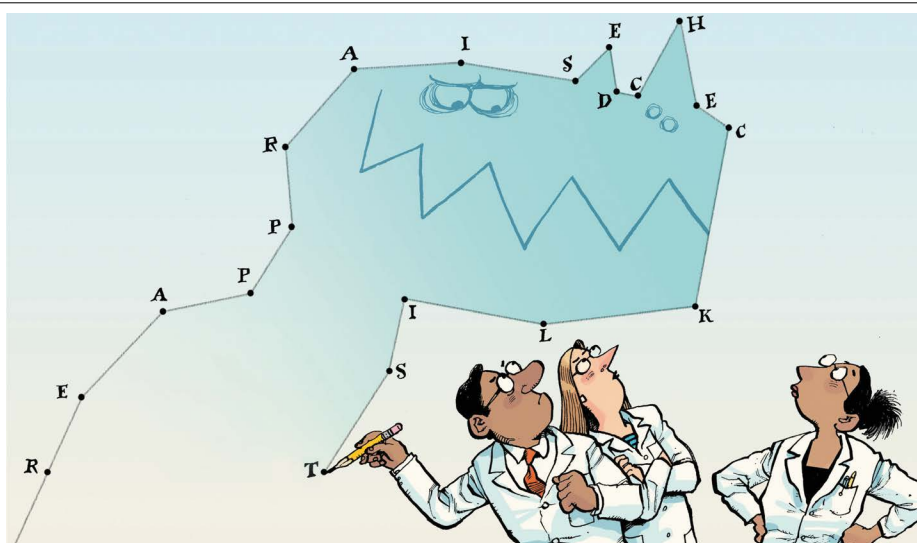
A tool that focuses on papers – not researcher behaviour – can help readers, editors and institutions assess which publications to trust.

If it is published in the scientific literature, can you trust it? All too often, that question gets lost, sidetracked or buried. Even when serious, credible concerns are sent to a journal, decisions over whether to correct or retract are more likely to take years than months – time during which potentially harmful misinformation can spread. Delays and inaction often happen because enquiries tend to focus on the thorny question of whether a researcher acted deliberately to deceive. The more important issue, however, is the integrity of the actual publication: is the research article reliable and are its conclusions valid?

As researchers who have spent years enmeshed in investigations, we think a major obstacle to evaluating the integrity of publications is a lack of tools. The Committee on Publication Ethics (COPE) advises publishers to retract articles when there is “clear evidence that the findings are unreliable”, but does not advise on how to determine whether that is the case. Resources for editors also focus on how to manage communications, rather than on how to evaluate reliability and validity. The net effect is inaction: readers remain uninformed about potential problems with a paper, and that can lead to wasted time and resources, and sometimes put patients at risk.

The integrity of a publication can be compromised in many ways. Some are unintentional: typos, transcription errors or incorrect analyses. Others are deliberate: image manipulation, data falsification and plagiarism¹. How publication integrity was compromised is secondary to whether the paper is reliable. Unreliable data or conclusions are problems irrespective of the cause.

Enter checklists, which have helped in the



structuring of complex procedures in health care and other industries. One of us (C.K.G.) helped to develop a checklist for assessing the quality of institutional investigations of researchers' conduct². Academic publishing has also introduced a suite of checklists to be completed by authors when submitting papers, to make sure certain aspects of the research are fully reported.

Here we present a tool – the REAPPRAISED checklist – that aims to help readers, journal editors and anyone else assess whether a paper has flaws that call its integrity into question. We developed it on the basis of our own experience and extensive consultation with research administrators and journal editors. Although designed for clinical and animal studies, the structured approach to investigation applies more broadly. Readers, peer reviewers, journals, publishers and institutions can use it to assess whether to trust a paper's findings. Our checklist should not be confused with journals' submission checklists, which are filled in by authors before publication and indicate what items are reported in the manuscript. The REAPPRAISED checklist can be used by anyone struggling to assess a submitted or published article, and includes common-sense assessments that go beyond the text itself. It can, and

should, be applied independently of whether misconduct is suspected. Its use can help to speed up the identification and correction of flawed papers, preventing wasted resources and even protecting patients from harm.

Cause for alarm

How did we come to see the need for this tool? From early 2013, three of us (A.A., A.G., M.J.B.) began to contact journals about multiple, serious problems we had identified in 33 reports of trials led by bone-health researchers Yoshihiro Sato and Jun Iwamoto³. The first retraction did not appear until late 2015.

This delay is all the more regrettable given that concerns had been raised more than a decade earlier. In 2003, a publication by Sato⁴ assessed a very rare complication of treatment of Parkinson's disease. Within a year, a letter⁵ to the editor remarked on how surprising it was that the research group had managed to identify 40 people with this complication in a very short time, because the writer's specialist institute had seen only two cases “in living memory”. Others had brought up concerns about ethical oversight⁶, as well as failure of randomization, implausible recruitment and outcomes, and other problems⁷, but no editorial comment or correction occurred at that time; retraction

Comment

THE 'REAPPRAISED' CHECKLIST FOR EVALUATION OF PUBLICATION INTEGRITY

Not all items will be applicable to every publication, and other questions might be relevant for individual categories.

R — Research governance

- Are the locations where the research took place specified, and is this information plausible?
- Is a funding source reported?
- Has the study been registered?
- Are details such as dates and study methods in the publication consistent with those in the registration documents?

E — Ethics

- Is there evidence that the work has been approved by a specific, recognized committee?
- Are there any concerns about unethical practice?

A — Authorship

- Do all authors meet criteria for authorship?
- Are contributorship statements present?
- Are contributorship statements complete?
- Is authorship of related papers consistent?
- Can co-authors attest to the reliability of the paper?

P — Productivity

- Is the volume of work reported by research group plausible, including that indicated by concurrent studies from the same group?
- Is the reported staffing adequate for the study conduct as reported?

P — Plagiarism

- Is there evidence of copied work?
- Is there evidence of text recycling (cutting and pasting text between papers), including text that is inconsistent with the study?

R — Research conduct

- Is the recruitment of participants plausible within the stated time frame for the research?
- Is the recruitment of participants plausible considering the epidemiology of the disease in the area of the study location?
- Do the numbers of animals purchased and housed align with numbers in the publication?
- Is the number of participant withdrawals compatible with the disease, age and timeline?
- Is the number of participant deaths compatible with the disease, age and timeline?
- Is the interval between study completion and manuscript submission plausible?
- Could the study plausibly be completed as described?

A — Analyses and methods

- Are the study methods plausible, at the location specified?
- Have the correct analyses been undertaken and reported?
- Is there evidence of poor methodology, including:
 - Missing data
 - Inappropriate data handling

- 'P-hacking': biased or selective analyses that promote fragile results
- Other unacknowledged multiple statistical testing
- Is there outcome switching — that is, do the analysis and discussion focus on measures other than those specified in registered analysis plans?

I — Image manipulation

- Is there evidence of manipulation or duplication of images?

S — Statistics and data

- Are any data impossible?
 - Are subgroup means incompatible with those for the whole cohort?
 - Are the reported summary data compatible with the reported range?
 - Are the summary outcome data identical across study groups?
 - Are there any discrepancies between data reported in figures, tables and text?
 - Are statistical test results compatible with reported data?
- Are any data implausible?
 - Are any of the baseline data excessively similar or different between randomized groups?
 - Are any of the outcome data unexpected outliers?
 - Are the frequencies of the outcomes unusual?
 - Are any data outside the expected range for sex, age or disease?
 - Are there any discrepancies between the values for percentage and absolute change?
 - Are there any discrepancies between reported data and participant inclusion criteria?
 - Are the variances in biological variables surprisingly consistent over time?

E — Errors

- Are correct units reported?
- Are numbers of participants correct and consistent throughout the publication?
- Are calculations of proportions and percentages correct?
- Are results internally consistent?
- Are the results of statistical testing internally consistent and plausible?
- Are other data errors present?
- Are there typographical errors?

D — Data duplication and reporting

- Have the data been published elsewhere?
- Is any duplicate reporting acknowledged or explained?
- How many data are duplicate reported?
- Are duplicate-reported data consistent between publications?
- Are relevant methods consistent between publications?
- Is there evidence of duplication of figures?

notices were finally issued in 2015–18.

Even after Sato admitted⁸ in 2016 to making up data, only 2 of the 34 journals we contacted took the initiative to assess other papers they had published by this group.

Investigations conducted at the four institutions at which Sato and Iwamoto worked were also, in our opinion, misdirected and incomplete. They focused on identifying researcher misconduct, not publication reliability. They assessed only 90 of 351 potentially compromised publications, and only after we contacted them in 2017 with concerns⁹. Two institutions failed to reach a conclusion about the reliability of 56 publications of the 78 they considered, because they could not determine whether misconduct had occurred. The other two did not report assessments on individual papers. (The institutions maintain that their investigations were appropriate, but have not responded directly to the criticisms, and two stated that it was difficult to assess publication integrity independently of misconduct.)

So far, at least 90 papers by Sato or Iwamoto have been retracted. More than twice that number remain in the literature, including 5 of the 33 clinical-trial reports. Readers have no way of knowing whether those reports are trustworthy. Nor do they have any guidance for making their own assessments.

These papers continue to be influential. In the years since we brought our concerns to journals, those 33 clinical-trial reports have been cited more than 600 times. They have been used by other researchers as evidence to help justify at least eight clinical trials¹⁰. Others that were later retracted were used in an effectiveness review conducted by the US Agency for Healthcare Research and Quality. They provided the sole evidence that bisphosphonates, commonly used for osteoporosis, could prevent fractures in patients at high risk of falls¹⁰.

Other cases had similar outcomes. More than ten years elapsed between the initial notification of concerns about publications of clinical studies in anaesthesia by Yoshitaka Fujii and the first of 183 retractions (see go.nature.com/2pvw2ax). Concerns raised about problems in a paper by Andrew Wakefield on autism and vaccination did not result in full retraction until 12 years after its publication. Flawed papers such as these influence future research initiatives, are incorporated into clinical guidelines and influence medical practice and public perceptions, so could potentially harm millions of patients.

Improving assessment

Our REAPPRAISED checklist facilitates systematic evaluation through 11 categories (see ‘The ‘REAPPRAISED’ list for evaluation of publication integrity’). It covers ethical oversight and funding, research productivity and investigator workload, validity of randomization, plausibility of results and duplicate

data reporting. It can identify problems from isolated data errors to data fabrication or falsification. Some of these questions should ideally have been asked by reviewers and editors before publication, and not all questions will apply to every paper, but it is still useful to consider all the questions collectively when assessing an article.

We developed the checklist in our roles as researchers (A.G., M.J.B., A.A.) and journal editor (A.A.K.) while evaluating thousands of publications; C.K.G., a specialist in institutional investigations, helped to refine it.

Our experience suggests that using REAPPRAISED can assist in decision-making before and after publication. A.A.K. implemented an earlier version of the checklist as editor-in-chief of *Anaesthesia*, which reviews nearly 1,000 submissions per year. Each is screened with the checklist; if concerns arise, individual patient data are requested and reviewed carefully for errors, inconsistencies or other red flags. In the past two years of routine use, editors have identified integrity problems in 42 submissions. A large subset of these – comprising work from 12 research groups – was serious enough to notify authors’ institutions. Six of the 12 ensuing investigations confirmed that data had been falsified or fabricated; two laid the blame on errors; and four are ongoing (A.A.K., personal communication).

We have also found this checklist effective in communicating with journals, having used it to submit a structured list of concerns to journal editors for the 56 publications for which institutional investigations were inconclusive. That has led to 29 retractions and 5 expressions of concern.

Implementation

We would like to see the checklist used during both manuscript review and post-publication evaluation. The fact that it separates assessment of publication integrity from the investigation of research misconduct will speed up evaluations. It could even be published alongside decisions to retract, issue an expression of concern, correct a paper, or let it stand.

We expect that use of REAPPRAISED will lead to more detailed, efficient, consistent and transparent evaluations of publication integrity, and thus faster and more accurate reporting of corrections and retractions. These improvements will benefit the researchers, clinicians, policymakers, patients and others who rely on the literature to make decisions. People using the tool will be able to help refine it as they gain experience, and it will help them to develop standards to assess the integrity of publications and act accordingly. We hope others will join in our efforts to implement and refine REAPPRAISED both informally and in future publications.

Integrity problems often cluster. Authors who have a paper retracted for misconduct are

more likely to accrue multiple retractions than are those whose work was retracted for other reasons¹¹. Publications by their co-authors can also be compromised¹². The checklist could help journal staff and investigation committees to decide when an assessment should be broadened to include other papers by a particular researcher and collaborators.

Ironically, a checklist that puts aside the question of misconduct might aid in evaluations of inappropriate behaviour. If multiple concerns are identified, or the concerns identified are those often associated with misconduct, the entire body of an author’s work should be systematically assessed.

Publishers’ integrity groups should adopt the checklist (or ask those expressing concerns to do so). Funders and government regulators should disseminate it to publishers, research institutions and other stakeholders. Peer reviewers and readers can use it on their own initiative, and those who have a nagging feeling about a publication can use it to work through their concerns and, if merited, communicate those to a journal.

If the goal is trustworthy literature, the integrity of publications – not just determination of misconduct – should be the focus of investigation.

The authors

Andrew Grey and **Mark J. Bolland** are associate professors of medicine at the University of Auckland, New Zealand. **Alison Avenell** is a professor at the Health Services Research Unit, University of Aberdeen, UK. **Andrew A. Klein** is a consultant anaesthetist at Royal Papworth Hospital, Cambridge, UK, and editor-in-chief of *Anaesthesia*. **C. K. Gunsalus** is the director of the National Center for Professional and Research Ethics at the University of Illinois at Urbana–Champaign, USA.
e-mail: a.grey@auckland.ac.nz

1. Nylenna, M. & Simonsen, S. *Lancet* **367**, 1882–1884 (2006).
2. Gunsalus, C. K., Marcus, A. R. & Oransky, I. *J. Am. Med. Assoc.* **319**, 1315–1316 (2018).
3. Bolland, M. J., Avenell, A., Gamble, G. D. & Grey, A. *Neurology* **87**, 2391–2402 (2016).
4. Sato, Y., Asoh, T., Metoki, N. & Satoh, K. *J. Neurol. Neurosurg. Psychiatry* **74**, 574–576 (2003); retraction **89**, e3 (2018).
5. Clarke, C. E. *J. Neurol. Neurosurg. Psychiatry* **75**, 510–511 (2004).
6. Caffrey, N., Casey, M. & Walsh, B. *J. Bone Miner. Res.* **21**, 491 (2006); correction **32**, 1151–1151 (2017).
7. Halbekath, J. M., Schenk, S., von Maxen, A., Meyer, G. & Mühlhauser, I. *Arch. Intern. Med.* **167**, 513–514 (2007).
8. Gross, R. A. *Neurology* **87**, 2388 (2016).
9. Grey, A., Bolland, M., Gamble, G. & Avenell, A. *Res. Integr. Peer Rev.* **4**, 3 (2019).
10. Avenell, A., Stewart, F., Grey, A., Gamble, G. & Bolland, M. *J. BMJ Open* **9**, e031909 (2019); correction **9**, e031909corr1 (2019).
11. Steen, R. G., Casadevall, A. & Fang, F. C. *PLoS ONE* **8**, e68397 (2013); correction <https://doi.org/10.1371/journal.pone.0068397> (2013).
12. Carlisle, J. B. & Loadman, J. A. *Anaesthesia* **72**, 17–27 (2017).