Assessing Research Misconduct in Randomized Controlled Trials

Wentao Li, MD, PhD, Esmée M. Bordewijk, MSc, and Ben W. Mol, FRANZCOG, PhD

Randomized controlled trials (RCTs) serve as the pillar of evidence-based medicine and guide medical practice. Compromised data integrity in RCTs undermines the authority of this valuable tool for science and puts patients at risk. Although a large number of retractions due to data issues in obstetrics and gynecology have occurred in the past few years, many problematic RCTs could still go uncovered because in general there is insufficient willingness to envisage and confront research misconduct. In this article, we discuss the necessity of assessing research misconduct, summarize methods that have been applied in detecting previous cases of misconduct, and propose potential solutions. There is no established mechanism to monitor feedback on published articles and the current system that handles potential research misconduct is unsatisfactory. Fortunately, there are methods to assess data integrity in RCTs both with and without individual participant data. Investigations into research misconduct can be facilitated by assessing all publications from a leading author or author group to identify duplication and patterns of ongoing misconduct. There is a pressing need to improve the mechanism that investigates data manipulation. The mechanism that handles misconduct should prioritize the interests of patients and readers rather than trial authors and their institutions. An equally urgent issue is to establish mechanisms that prevent compromised trials from polluting evidence synthesis or misleading practice.

(Obstet Gynecol 2021;138:338–47)
DOI: 10.1097/AOG.0000000000004513
notices and information on the publisher’s website frequently used euphemistic language and concealed reasons for retraction.7

Here, we discuss the scale of the problem, summarize methods that have been applied in detecting previous cases of misconduct, and propose potential solutions.

SCOPE OF THE PROBLEM AND BARRIERS TO OVERCOME

We performed a search of retraction notice or expression of concern for RCTs in obstetrics and gynecology in the Retraction Watch database, where the reasons for rejections were not only assessed from the notices of publishers but also coupled with the information that Retraction Watch gained independently.8 The search, performed on December 22, 2020, was limited to “(HSC) Medicine—Obstetrics or Gynaecology” for the subject and “Clinical Study” for article type. No time restriction was imposed on the search. We excluded notices not related to RCTs, obstetrics and gynecology (not including breast-related studies), or data validity. In cases where an expression of concern was followed by a retraction notice, only the retraction notice was retained to avoid duplication. We added several recent notices that we were aware of but have not been included in the database.

A total of 36 RCTs in obstetrics and gynecology have been retracted or are undergoing formal investigations due to concerns related to data validity (Appendix 1, available online at http://links.lww.com/AOG/C389). A number of reputable journals in obstetrics and gynecology published these manuscripts. Studies from Japan (n = 16) and Egypt (n = 15) account for 86% of the notices. About 81% of identified studies were published by six first authors who had multiple retractions or expressions of concern. One third of these articles were published before 2005, when precautionary measures such as trial registration have not been widely deployed by journals until the recent decade. There is no consistent protocol nor motivation to monitor feedback on articles once they are published, let alone the enormous resources needed to audit published trials.

In reality, the decision to retract an article is not taken lightly by journals. On average, it takes almost 4 years for articles labeled with research misconduct to be retracted across all fields of medicine.11 Judging from the much longer duration (11.2 years) from publication to retraction for RCTs with data issues in obstetrics and gynecology, these barriers could be more difficult to overcome in our field.

Publication is usually viewed as the endpoint of research projects. Coherent mechanisms to monitor feedback of readers do not exist. The attitude of publishers and journals toward warnings about research misconduct varies significantly.3 Some respond seriously and act in a timely fashion, but many are slow or tend to ignore allegations of misconduct. Lack of willingness to envisage integrity issues of published articles often prolongs or prevents retraction of problematic articles. Reluctance to receive warnings can also frustrate and discourage whistle-blowers and sometimes puts them at risk, such as being retaliated on by the authors.

Publishers and journals are committed to retraction guidelines of COPE (the Committee on Publication Ethics) to handle misconduct issues.12 Although the COPE guidelines provide basic guidance for handling concerns regarding the integrity of publications, they should not be viewed as perfect tools to solve these concerns. The COPE guidelines follow a rigid branching logic and rely heavily on responses from authors and their institutions. Any hiccups in communication could cause months or years of delay in the investigation. As one might imagine, in most investigations of research misconduct, neither authors nor their institutions respond adequately in a timely manner. Given the usual absence of regulatory oversight, the motivation of institutions to investigate the research of their employees is largely questionable.

Another concern is that the COPE guidelines, although endorsed by most biomedical journals, are not implemented consistently. In cases of the most prolific authors who committed fraud, involved journals displayed huge variations in practice when
mounting evidence confirmed massive fraud. It took more than 5 years for some articles in the series to be retracted.3

The above-mentioned issues in the current system form a black box that hinders general readers from knowing investigations about the evidence they use until these articles are retracted suddenly, usually years after publication. Independent assessments could be powerful supplements to formal investigations that have a high threshold to initiate. However, few people are equipped with adequate knowledge to detect data integrity issues in RCTs and there is no overview of available methods that disclose these problems. These barriers need to be removed to invite more interest in evaluating manuscripts for research misconduct.

METHODS TO ASSESS RESEARCH MISCONDUCT IN RANDOMIZED CONTROLLED TRIALS

Research misconduct is commonly defined as, “fabrication, falsification, or plagiarism in proposing, performing, or reviewing research, or in reporting research results.”13 Here we discuss only fabrication and falsification, because plagiarism has attracted wide attention and effective tools have been implemented to detect plagiarism. Fabrication refers to making up data, results, or recordings and reporting them. Falsification means manipulating research materials, equipment, or processes or changing or omitting data or results such that the research is not accurately represented in the research record.

When assessing data integrity, comparing all publications of the leading author or author group being investigated may offer extra insights into the problem because certain patterns of data manipulation that are unique to the leading author or group can only be identified when all their works are analyzed and compared. As an example, we examined all the RCTs of two authors from the same university in the field of obstetrics and gynecology.14 Using pairwise comparisons of baseline characteristics and outcome tables, for 21 of the 35 evaluated trials, more than 10 identical or extremely similar values between trials were observed, uncovering a pattern of self-plagiarism of data. A short-fall of our approach is that this process becomes more demanding when the subject of the investigation has a large number of publications.

Screening for Integrity Issues

Most investigations regarding cases of compromised data integrity are triggered by suspicious clues of implausible study design, conduct, or report. This screening process to assess overall integrity is crucial to identifying targets of investigation.

Assessors need to consider a wide range of issues, including but not limited to ethics, research governance, authorship, productivity, analytical methods, effect direction and scale, consistency of data, reporting, and funding. Grey et al formalized the assessment of individual articles, not restricted to RCTs, as the “REAPPRAISED” checklist.15 This checklist facilitates systematic evaluation through 11 categories and, as the authors suggested, can be used during both manuscript review and postpublication evaluation. A pilot study was conducted in which this checklist was used as a screening tool for editorial works of one journal but no formal validation has been reported.

We propose a checklist to assess the integrity of RCTs that assisted in our investigations (Table 1).

Absent or retrospective registration suggests poor transparency of the trial plan. Similarly, absent or vague descriptions of research ethics may indicate that the trial lacks ethics approval. These issues are usually found in trials that were determined to be fabricated.

Reports of RCTs often involve a large author group because of the complexity of trial conduct. However, leading authors who commit fraud may wish to keep the author group small to minimize the risk of being exposed. Thus, a low author-to-study size ratio should trigger concern. Special attention should be paid to authors who were involved in retracted articles for misconduct or researchers who published a large number of RCTs in a short period as leading authors.

We also encountered scenarios where the plausibility of using placebo or intervention is in doubt; for example, in a three-arm placebo-controlled trial, two interventions were administered via different routes and look different, but there was only one placebo group to mask two entirely different interventions. Placebo preparation is technically challenging and usually requires the involvement of manufacturers of interventions. It is unlikely that a placebo-controlled trial can be performed without funding or the involvement of manufacturers.

Authors who fabricate trials may report a trial that accumulated a large number of participants in a short period that exceeded the capacity of their institutions. They may present a fast submission that ignores the time needed to finish follow-up, perform data analysis, and write the manuscript. Besides implausible time, we also noticed that fabricated trials frequently reported zero participants lost to follow-up or reported loss to follow-up leading to perfectly balanced numbers for analysis (eg, three groups of 50; two groups of 100).
Information in baseline characteristics and outcomes can offer insight into integrity. They should be assessed from a methodologic point of view such as extreme balance for baseline characteristics and consistency between outcomes. More importantly, judgement according to common sense, the literature, and local data can be extremely helpful. For example, when assessing obstetrics trials, one could assess whether summary values of gestational age and birthweight are reasonably correlated because researchers who commit fraud may ignore necessary correlations between these two related parameters.

We have not been able to determine the most appropriate threshold to trigger an investigation. In our experience, cases of misconduct are likely to meet several of the criteria in the checklist. However, some criteria have more weight than others; for example, any issue regarding author group, intervention, and timeframe could offer enough caution to question the veracity of the study. We encourage journal editors to pilot this checklist in editorial works and publicize findings to accumulate empirical evidence before a formal validation work with adequate cases and controls will be possible.

### Data Integrity Assessment Without Individual Participant Data

Although sharing of individual participant data has been encouraged for several years, data sharing is still uncommon for published RCTs. Several methods can be applied to assess data integrity using summary statistics in publications without individual participant data available (Table 2); however, abnormalities found with summary statistics could be explained by honest error, poor reporting, or misconduct.

Some reported statistical results can be reproduced with summary data. This approach is especially suitable for univariable parametric analyses, which are commonly used in RCTs. Independent t test and one-way analysis of variance can be reproduced with reported means, SDs, and the sample size reported in articles. Chi-square tests, Fisher exact tests, odds ratios, and risk ratios can be reproduced using absolute numbers given in crosstabs. Recalculated P-values can then be checked against the reported P-values for inconsistencies. A software called Statcheck can extract reported statistical values from the text to reproduce the P-value for the reported statistical result. However, this tool does not work for tables and may not

---

**Table 1. Proposed Checklist for the Integrity of Randomized Controlled Trials**

<table>
<thead>
<tr>
<th>Domain</th>
<th>Item</th>
</tr>
</thead>
<tbody>
<tr>
<td>Governance</td>
<td>Absent or retrospective registration for RCTs; this is more relevant in RCTs started after 2010</td>
</tr>
<tr>
<td></td>
<td>Absent or vague description of research ethics or apparent deviation from basic principles of ethical research even though ethics approval was claimed</td>
</tr>
<tr>
<td></td>
<td>Mismatch between the RCT report and trial registration, especially design, sample size, and time</td>
</tr>
<tr>
<td>Author group</td>
<td>No. of authors less than 3 or low author/study size ratio, especially RCTs with 1 or 2 authors only</td>
</tr>
<tr>
<td></td>
<td>Other studies of the first author or co-authors being retracted for misconduct (search PubMed and, if needed, Retraction Watch)</td>
</tr>
<tr>
<td></td>
<td>Large number of RCTs published in a short period by one author or in one institute (eg, more than 3 per year as first author)</td>
</tr>
<tr>
<td>Plausibility of intervention</td>
<td>Implausible use of placebo or intervention (eg, 2 interventions with 1 placebo)</td>
</tr>
<tr>
<td>Timeframe</td>
<td>Implausibly fast recruitment of participants considering the prevalence of the condition, capacity of recruitment center, and eligibility criteria</td>
</tr>
<tr>
<td></td>
<td>Implausible amount of time between ending of recruitment, ending of follow-up, and submission of the paper (take into account time between randomization and time to assessment of the primary report)</td>
</tr>
<tr>
<td>Drop-out rates</td>
<td>Zero participants lost to follow-up or no information on loss to follow-up</td>
</tr>
<tr>
<td>Baseline characteristics</td>
<td>Loss to follow-up leads to perfectly rounded numbers for analysis (eg, 3 groups of 50, 2 groups of 100)</td>
</tr>
<tr>
<td></td>
<td>No or few baseline characteristics presented</td>
</tr>
<tr>
<td></td>
<td>Implausible patient characteristics judging from common sense, the literature, and local data (eg, SDs are similar for completely different characteristics with different means or distributions)</td>
</tr>
<tr>
<td></td>
<td>Extreme balance for baseline characteristics between arms that is not compatible with chance</td>
</tr>
<tr>
<td>Outcomes</td>
<td>Effect size that was much larger than in other RCTs regarding the same topic (heterogeneity)</td>
</tr>
<tr>
<td></td>
<td>Highly significant results with a small sample size</td>
</tr>
<tr>
<td></td>
<td>Conflicting information between outcomes (eg, more ongoing pregnancies than clinical pregnancies)</td>
</tr>
<tr>
<td></td>
<td>Recalculated P-values are inconsistent with the reported P-values</td>
</tr>
</tbody>
</table>

RCT, randomized controlled trial.
be compatible with formats other than the American Psychological Association (APA).

The GRIMMER (Granularity-Related Inconsistency of Means Mapped to Error Repeats) test allows for testing whether reported measures of variability are possible by assessing whether a reported statistic is consistent with the sample size and granularity.20,21 This method can test variances, SDs, and standard errors. It is possible to extend the test to other measures of variability or apply the test to noninteger data. A limitation of the software is that it only works for a sample size of 99 or less.

SPRITE (Sample Parameter Reconstruction via Iterative Techniques) is a technique for reconstructing potential discrete data sets using basic summary information about a sample, namely the mean, the SD, the sample size, and the lower and upper bounds of the range of item values.22 SPRITE complements GRIMMER tests. It does not have a sample size limitation and can be used with per-cell sample sizes in the thousands. SPRITE also takes into consideration the range of possible values of the raw data.

Benford’s law describes the distribution of the digits in naturally occurring numbers.23 The first digits tend to follow a special distribution where digits 1, 2, and 3 account for more than 60% of the total probability distribution. The last digits, however, approach uniform distribution. It is possible to extract numbers in one article and assess whether the frequencies of the first digits and last digits are consistent with Benford’s law.23,24 An excessive proportion of certain digits might indicate that the authors unconsciously prefer these digits when making up numbers. However, not all numbers in research occur naturally and some naturally occurring numbers may have unequal distributions of digits; for example, laboratory readings might be rounded. A validation study using cases of fraud and nonverified controls found that applying Benford’s law to RCTs is unlikely to miss cases (100% sensitivity) but may misclassify many controls as cases (poor specificity).23

### Table 2. Applicability of Methods That Assess Data Integrity in Randomized Controlled Trials

<table>
<thead>
<tr>
<th>Method</th>
<th>Minimum Information Needed</th>
<th>Link to Resource or Program</th>
</tr>
</thead>
<tbody>
<tr>
<td>Screening for integrity: the REAPPRAISED checklist</td>
<td>Manuscript+tables+figures</td>
<td><a href="http://resource-cms.springernature.com/springer-cms/rest/v1/content/17589730/data/v1">http://resource-cms.springernature.com/springer-cms/rest/v1/content/17589730/data/v1</a></td>
</tr>
<tr>
<td>Data integrity</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Reproduce univariable P-value or effect estimates with summary statistics</td>
<td>Manuscript+tables</td>
<td><a href="https://cran.r-project.org/web/packages/statcheck/index.html">https://cran.r-project.org/web/packages/statcheck/index.html</a> <a href="https://cran.r-project.org/web/packages/rpsychi/index.html">https://cran.r-project.org/web/packages/rpsychi/index.html</a></td>
</tr>
<tr>
<td>GRIMMER test</td>
<td>Manuscript+tables</td>
<td><a href="http://www.prepubmed.org/grimmer/">http://www.prepubmed.org/grimmer/</a></td>
</tr>
<tr>
<td>SPRITE</td>
<td>Manuscript+tables</td>
<td><a href="http://www.prepubmed.org/sprite/">http://www.prepubmed.org/sprite/</a></td>
</tr>
<tr>
<td>Digit preference (Benford’s law)</td>
<td>Manuscript+tables+figures</td>
<td><a href="http://www.ctc.ucl.ac.uk/Training.aspx">http://www.ctc.ucl.ac.uk/Training.aspx</a> <a href="https://github.com/chartgerink/ddfab">https://github.com/chartgerink/ddfab</a></td>
</tr>
<tr>
<td>Baseline P-value distribution for RCTs*</td>
<td>Manuscript+baseline table</td>
<td><a href="https://cran.r-project.org/web/packages/simdist/index.html">https://cran.r-project.org/web/packages/simdist/index.html</a></td>
</tr>
<tr>
<td>Reproduce P-value or effect estimates with IPD</td>
<td>Manuscript+tables+IPD</td>
<td>NA</td>
</tr>
<tr>
<td>Pattern in repeated measurements</td>
<td>Manuscript+tables+IPD</td>
<td><a href="http://www.ctc.ucl.ac.uk/Training.aspx">http://www.ctc.ucl.ac.uk/Training.aspx</a></td>
</tr>
<tr>
<td>Pattern of missing values</td>
<td>Manuscript+tables+IPD</td>
<td>R-program available of the Web Appendix of <a href="https://doi.org/10.1016/j.jclinepi.2017.03.018">https://doi.org/10.1016/j.jclinepi.2017.03.018</a></td>
</tr>
<tr>
<td>Unusual sequences after sorting by allocation or outcomes</td>
<td>Manuscript+tables+IPD</td>
<td>NA</td>
</tr>
<tr>
<td>Rationality of dates for events, recruitment, and visits</td>
<td>Manuscript+tables+IPD</td>
<td><a href="http://www.ctc.ucl.ac.uk/Training.aspx">http://www.ctc.ucl.ac.uk/Training.aspx</a></td>
</tr>
<tr>
<td>Recruitment over time by allocation</td>
<td>Manuscript+tables</td>
<td>R-program available of the Web Appendix of <a href="https://doi.org/10.1016/j.jclinepi.2017.03.018">https://doi.org/10.1016/j.jclinepi.2017.03.018</a></td>
</tr>
<tr>
<td>Strength and direction of correlations between variables</td>
<td>Manuscript+tables</td>
<td><a href="http://www.ctc.ucl.ac.uk/Training.aspx">http://www.ctc.ucl.ac.uk/Training.aspx</a></td>
</tr>
<tr>
<td>Coefficients of variation for unrelated variables</td>
<td>Manuscript+tables</td>
<td>NA</td>
</tr>
<tr>
<td>Inliers</td>
<td>Manuscript+tables+IPD</td>
<td><a href="http://www.ctc.ucl.ac.uk/Training.aspx">http://www.ctc.ucl.ac.uk/Training.aspx</a></td>
</tr>
<tr>
<td>CSM (for multicenter trial only)</td>
<td>Manuscript+tables+IPD</td>
<td><a href="http://www.ctc.ucl.ac.uk/Training.aspx">http://www.ctc.ucl.ac.uk/Training.aspx</a></td>
</tr>
</tbody>
</table>

IPD, individual participant data; NA, not available; GRIMMER, Granularity-Related Inconsistency of Means Mapped to Error Repeats; SPRITE, Sample Parameter Reconstruction via Iterative Techniques; CSM, central statistical monitoring.

* May need supplementary materials.
There is a common misconception that \( P \) values generated from statistical tests that compare baseline characteristics should all be greater than 0.05 to demonstrate the success of randomization. In fact, \( P \) values in the baseline characteristics table are expected to approach a uniform distribution between 0 and 1. This balance of characteristics table are expected to approach a uniform distribution. Further works were also performed to estimate the expected distributions using rounded summary statistics of a large collection of authentic RCTs.

**Data Integrity Assessment With Individual Participant Data**

The assessment of data integrity with individual participant data is more likely to provide solid evidence of misconduct compared with an evaluation of summary statistics because it benefits from the abundance and complexity of information hidden in the data sets. However, the biggest hurdle for these assessments is the unavailability of data for submitted or published trials.

Reproducing statistics using individual participant data is an indispensable step when assessing data integrity. Some patterns of data manipulation can be easily identified without any calculations. For example, in trials with repeated measurements of outcomes on the same individuals, propagation of previous data to later data leads to insufficient variability between consecutive observations. By sorting observations according to treatment allocation or outcomes, unusual sequences such as repeated values or ordered progressions may be evident. Also, fabricated data tend to be perfect in terms of no or few missing values, whereas, in real trials, 100% follow-up is extremely rare.

The rationality of dates may be overlooked by authors when fabricating data. For example, no dates should occur before the first recruitment, and, similarly, no events should happen after death or the end of the study. It might be possible to check whether there are too many participant visits that took place during weekends or public holidays. Also, the rate of real participant recruitment or randomization is unlikely to be perfectly constant over time as recruitment in trials usually have a “climbing” period. Performing a comparison of recruitment speed alongside time between treatment groups may reveal periods during which treatment allocation was manipulated.

It is difficult to fabricate several variables that are correlated and be consistent with real data. Some variables should be correlated based on knowledge or common sense; for example, the volume of blood loss and drop in hemoglobin should be related in a trial studying postpartum hemorrhage and, the strength of correlation should be consistent across treatment groups. When falsifying these data, the correlations may end up too strong or too weak to be plausible. It is also difficult to fabricate several means and SDs for separate variables or groups in a way that emulates real data. Authors who commit fraud may come up with variables that are too similar for unrelated variables. This can be assessed with coefficients of variation.

Inliers that refer to strange distributions of values that are too close to the overall study mean indicate concerning regards data integrity. When manipulating data, people tend to avoid values that are far from the mean as they are more noticeable. There is also a popular misconception that large SDs, irrespective of sample size, indicate unsatisfactory sampling. To avoid outliers, they may create a large number of inliers that lead to very small variations.

Central statistical monitoring refers to comparing data from one center with data from other centers using statistical methods. It is possible to apply this method in the assessment of data integrity in a multicenter trial. This method is based on an important assumption that anyone who manipulated data did not have access to trial data from other staff or centers. As a result, manipulated data might present different characteristics from true observations. Statistical aspects such as the missing value pattern, distribution of values, correlation between variables, and inliers can be compared between centers. Visualizing these differences with graphs could facilitate the assessments.

**SAFEGUARDING THE INTEGRITY OF TRIALS**

Publications containing falsified data not only harm the reputation of journals and publishers but also misguide future research and put patients at risk. Everyone in the research community should take a firm stand to oppose data manipulation in research and support the efforts to denounce such behavior. We summarize our proposed action points in Box 1.
The mechanism investigating data integrity in RCTs needs to prioritize the interests of patients and readers rather than trial authors. Improved efficiency, consistency, and transparency should be stressed when arguing that biomedical research centers on benefitting patients.

Measures to safeguard the trustworthiness of RCTs before publication, mainly trial registration and the CONSORT (Consolidated Standards of Reporting Trials) statement, are endorsed by most journals. The latter, however, aims to promote transparent reporting of trial methodology, which is different from the detection of data fabrication. More importantly, their actual implementation may not be satisfactory. The requirement of compulsory prospective registration for any clinical trials was launched 17 years ago. However, even high-impact journals continue to publish trials that do not meet this standard. We previously demonstrated that RCTs that clearly violated the CONSORT statement can be published in peer-reviewed journals. A more concerning issue is that there have been examples where retracted RCTs due to data integrity were prospectively registered, which means the rule of registration has been deliberately bypassed by some researchers. On the other hand, trial registration has been extremely violated, which means the rule of registration has been deliberately bypassed by some researchers. On the other hand, trial registration has been extremely helpful in the detection of fabrication, as inconsistencies regarding the start and end date of studies and sample size help to uncover fraud.

At present, journals’ retraction practices are inconsistent and many retractions fail to state the reason for retraction. This explains a paradoxical phenomenon that some retracted articles were cited far more often after their retractions. Also, the COPE retraction guidelines may consider parallel routes of action where expressions of concern are instantly issued when investigations embark. Further elaboration in the COPE guidelines of the threshold to initiate investigation, maximum waiting time for responses, and the format of retraction notices would be useful.

Information sharing regarding research misconduct between journals should be smoother to align with the changing publishing environment. Journals do not share information of submissions with their counterparts, which makes it easy for authors who commit research misconduct to walk away from a bad record and submit the compromised paper to another journal. Dedicated channels could be established to facilitate the sharing of information about research misconduct.

Understandably, journals and editors may have limited energy and resources to oversee the integrity of every submission. Peer-review, a mechanism to assess scientific values of submissions, was not designed to identify integrity problems. Thus, independent research and open discussion about research misconduct should be encouraged as this issue can only be deterred by joint efforts and common awareness. Currently, readers post their doubts at publications review websites or public platforms such as

### Box 1. Proposed Actions to Improve Integrity in Randomized Controlled Trials

#### Journals and Publishers
- Strictly abide by the requirements of prospective registration and CONSORT statement for publication
- Consider the possibility of research ethics breach and encourage relevant assessments
- Share information regarding research misconduct with other journals
- Set up platforms to communicate with readers and respond to whistleblowers in a timely and serious manner
- Involve independent investigators or experts in investigations
- Issue expression of concern when investigations take place
- Provide explicit reasons for retractions

#### COPE (Committee on Publication Ethics)
- Define the threshold to initiate an investigation, maximum waiting time, and format of retraction notice
- Consider alternative routes in the case of slow or no response from authors or institutions

#### Institutions and Regulatory Bodies
- Provide adequate training regarding research ethics to researchers
- Actively monitor research integrity of employees
- Rationalize promotion criteria and avoid monetary rewards to researchers for publication
- Establish and execute penalty mechanisms for research misconduct
- Respond to inquiries from journals in a timely and serious manner and establish committees for investigation

#### Authors of Evidence Synthesis or Guidelines
- Use checklists to identify potential integrity problems for each included RCT
- Write to authors for clarification on concerns
- Include only trustworthy RCTs that pass assessments
- Move toward an individual participant data–based approach

#### Others
- Consider compulsory data sharing as a condition for publication of RCTs
- Encourage open discussion of research misconduct

CONSORT, Consolidated Standards of Reporting Trials; RCT, randomized controlled trial.
as Twitter, which may not necessarily catch the attention of journals and publishers. The gap between readers and journals could be mitigated by establishing dedicated platforms to report and act.

An equally urgent issue is to set up mechanisms that prevent compromised trials from being considered in evidence synthesis and clinical practice. The attempt to include all relevant studies to minimize the risk of publication bias in systematic reviews could be counterproductive because it facilitates contamination of evidence from unverified trials. There should be a thorough consideration regarding the validity of trials before pooling or use of trials for meta-analyses. For example, authors of evidence synthesis can use the RE-APPRAISED checklist or Table 1 to analyze the risk of individual trials and query authors of trials if there are any data validity concerns. Trials with significant concerns should not be included unless they were judged to be safe after clarification. We previously demonstrated that trials with individual participant data sharing have better performance on quality and integrity than trials without sharing. Individual participant data meta-analyses possess advantages in terms of generating more reliable evidence because those data included need to pass quality and integrity checks. Therefore, evidence synthesis should move toward an individual participant data-based approach. Unfortunately, these attempts are usually hampered by the unavailability of data at present. As an obligation of publication, a compulsory requirement to submit research data to appropriate data repositories for replication of results and other righteous purposes may be part of the solution.

CONCLUSIONS

Data fabrication or falsification in RCTs should not be overlooked. A growing number of methods can be applied to assess data integrity in RCTs. We call for regular assessments of data integrity of trials before their acceptance for publication or inclusion in evidence synthesis or guidelines.

REFERENCES

6. Chambers L, Michener C, Falcone T. Plagiarism and data falsification are the most common reasons for retracted publications in obstetrics and gynaecology. BJOG 2019;126:1134–40. doi: 10.1111/1471-0528.15689
7. Li W, Mol BW. Re: plagiarism and data falsification are the most common reasons for retracted publications in obstetrics and gynaecology. BJOG 2019;126:1289. doi: 10.1111/1471-0528.15829
tion via Iterative Techniques (SPRITE), PeerJ Preprints 2018;6: e26968v1. doi: 10.7287/peerj.preprints.26968v1


42. Retraction: nifedipine alone or combined with sildenafil citrate for management of threatened preterm labour: a randomised trial. BJOG 2021;128:144. doi: 10.1111/1471-0528.16565


PEER REVIEW HISTORY


CME FOR THE CLINICAL EXPERT SERIES

Learning Objectives for “Assessing Research Misconduct in Randomized Controlled Trials”

After completing this learning experience, the involved learner should be able to:

• Discuss how compromised data integrity in randomized controlled trials (RCTs) affects patient care;

• Summarize methods that have been applied in detecting previous cases of misconduct;

• Propose potential solutions to identify and callout research misconduct;

• List ways to improve the mechanism that investigates data manipulation.

Instructions for Obtaining AMA PRA Category 1 Credits™

Continuing Medical Education credit is provided through joint providershhip with The American College of Obstetricians and Gynecologists. Obstetrics & Gynecology includes CME-certified content that is designed to meet the educational needs of its readers. This article is

346 Li et al Misconduct in Randomized Controlled Trials

OBSTETRICS & GYNECOLOGY

© 2021 by the American College of Obstetricians and Gynecologists. Published by Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.
certified for 2 AMA PRA Category 1 Credits™ This activity is available for credit through September 30, 2024.

Accreditation Statement

ACCME Accreditation
The American College of Obstetricians and Gynecologists is accredited by the Accreditation Council for Continuing Medical Education (ACCME) to provide continuing medical education for physicians.

AMA PRA Category 1 Credit(s)™
The American College of Obstetricians and Gynecologists designates this journal-based CME activity for a maximum of 2 AMA PRA Category 1 Credits™. Physicians should claim only the credit commensurate with the extent of their participation in the activity.

College Cognate Credit(s)
The American College of Obstetricians and Gynecologists designates this journal-based CME activity for a maximum of 2 Category 1 College Cognate Credits. The College has a reciprocity agreement with the AMA that allows AMA PRA Category 1 Credits™ to be equivalent to College Cognate Credits.

Disclosure of Faculty and Planning Committee Industry Relationships
In accordance with the College policy, all faculty and planning committee members have signed a conflict of interest statement in which they have disclosed any financial interests or other relationships with industry relative to article topics. Such disclosures allow the participant to evaluate better the objectivity of the information presented in the articles.

How to Earn CME Credit
To earn CME credit, you must read the article in Obstetrics & Gynecology and complete the quiz, answering at least 70 percent of the questions correctly. For more information on this CME educational offering, visit the Lippincott CMEConnection portal at https://cme.lww.com/browse/sources/196 to register and to complete the CME activity online. ACOG Fellows will receive 50% off by using coupon code, ONG50.

Hardware/software requirements are a desktop or laptop computer (Mac or PC) and an Internet browser. This activity is available for credit through September 30, 2024. To receive proper credits for this activity, each participant will need to make sure that the information on their profile for the CME platform (where this activity is located) is updated with 1) their date of birth (month and day only) and 2) their ACOG ID. In addition, participants should select that they are board-certified in obstetrics and gynecology.

The privacy policies for the Obstetrics & Gynecology website and the Lippincott CMEConnection portal are available at http://www.greenjournal.org and https://cme.lww.com/browse/sources/196, respectively.

Contact Information
Questions related to transcripts may be directed to educationcme@acog.org. For other queries, please contact the Obstetrics & Gynecology Editorial Office, 202-314-2317 or obgyn@greenjournal.org. For queries related to the CME test online, please contact cconnection@wolterskluwer.com or 1-800-787-8985.

© 2021 by the American College of Obstetricians and Gynecologists. Published by Wolters Kluwer Health, Inc. Unauthorized reproduction of this article is prohibited.