

A Catalogue of Errors?

Nick Freemantle PhD, Professor of Clinical Epidemiology & Biostatistics, Research Department of Primary Care and Population Health, and Priment Clinical Trials Unit, University College London, London, UK.

Greta Rait MD, Senior Lecturer in Primary Care, Research Department of Primary Care and Population Health, and Priment Clinical Trials Unit, University College London, London, UK.

Address for Correspondence

Nick Freemantle PhD
Professor of Clinical Epidemiology & Biostatistics
Department of Primary Care and Population Health
PRIMENT Clinical Trials Unit
Upper Third Floor
UCL Medical School (Royal Free Campus)
Rowland Hill Street
London
NW3 2PF
Email: nicholas.freemantle@ucl.ac.uk
Tel: +44 (0)20 7794 0500 ex34756
Fax: +44 (0)20 7794 1224

The paper by Cole and colleagues [1] examines the relationship between discrepancies and retractions in clinical trial reports, concluding that discrepancies may be an early signal of unreliability. It is another foray by the authors in the examination of error in reported research. The previous paper [2,3] demonstrated a clear and concerning relationship between error and effect size in stem cell research. But is identifying fraudulent research in clinical trial reports really as simple as counting the number of errors?

Given the authors' broad definition of a clinical trial, we elected to apply the taxonomy that they recommend in an appraisal of their current paper.[1] We independently assessed the paper for errors, collated them and came to a consensus on the quality of the paper and its interpretation. In the following paragraphs we describe what we found then make some concluding comments.

First, the authors undertake a case control study and make recommendations based upon sensitivity and specificity, even though they acknowledge that these metrics are dependent upon prevalence and the research design is inappropriate for these methods.[4] The prevalence of unretracted clinical trial papers is very large (802953 on www.pubmed.org on 6th July 2015), and of retracted clinical trial papers is blindingly small (379 on same date), so the predictive power of an error-count to identify rogue papers is extremely low. This is illustrated by their own table of included studies for which 15 (30%) of papers in the retracted group have <3 errors (the level at which they suggest that a 'red flag' is raised) and 17 (33%) of papers in the nonretracted controls have ≥ 3 errors. You would have to kiss a lot of princes to find a frog.

Second, they interpret their outcome (retracted paper) as if it is always orthogonal, homogeneous, and harmful. A duplicate publication, while wrong, may paradoxically help the practitioner to identify an important piece of information that they might otherwise miss, but fabricated results may lead to patient harm. Their primary outcome cannot be considered: *'the variable capable of providing the most clinically relevant and convincing evidence directly related to the primary*

objective...' . [5] Instead they might have chosen a particular category of retracted paper, such as misconduct, or at least excluded the 12 (24%) duplicate and plagiarised ones.

Third, the authors blind the papers, but provide the Nat Library of Med ID, making it possible but tedious for those readers with access to the internet to identify the journals and scrutinise their study material. Basic demographics are an essential part of good reporting practice – in this case this should involve the journal and specialty. It is for example important that 8 (16%) retracted papers come from the same journal [6] and on the basis of the journal titles alone 15 (30%) are in the area of anaesthesia which, while important, is over represented in a way that requires further scrutiny. Indeed, assessing the retracted papers in anaesthesia we quickly discover evidence of systematic fraud [7] focused on the work of two authors who between them have had 102 retracted articles.

Fourth, the authors undertake an analysis of the relationship between errors and retraction in their case control study. Given the hierarchical nature of their data (the 5 journals with multiple papers which include 17 (34%) of papers, the limited number of clinical areas) they fail to account for the repeated journals, which is curious when elsewhere they recognise that papers from the same journal cannot be considered independent.

Fifth, the authors kept on looking for errors even though as individuals none of them appears to have found the majority. Did knowing that a paper had many errors lead them to look for more? How did they decide to stop looking? We looked independently and stopped quickly with 5 between us as identifying more seemed unnecessary.

So what should we make of the paper? The authors identify a relationship between errors in clinical trial reports and publication retraction which may be true but is not helpful in practice. Counting up errors could lead us to increase our suspicions that a paper may have a higher chance of being

retracted, which might be because it was clinically misleading but could also be because of another reason like duplication.

Fundamentally we should ask why these errors were not picked up by the editors or technical editors who should ensure the validity of the work they publish? Plagiarism has become much easier to identify using online tools, but it appears that misconduct may not be acted upon even when it is identified by a regulatory agency.[8]

Our experience is that many journals manage the publication process very well. It is surprising therefore that the paper with 35 errors (in a BMJ group journal) has 34 purported numerical errors. Schafer describes the considerable efforts that editors went to in uncovering the extent of research fraud in the work of authors in anaesthesia after errors were highlighted by readers [7], and recognises the duty of care of the journal to ensure high quality research. But he likened the growth of human knowledge to the weaving of a tapestry, and the sudden loss of retracted articles to the ripping of a thread. Many papers are published but fortunately few are fabricated or irrevocably flawed. Alexander Pope, in his *Essay on Criticism*, commented that: *'To err is Human'*. We must expect both benign and malicious errors to occur and devise ways to identify and address errors of substance in the research publication process. However, the authors' recommendation that we 'raise a red flag' for clinical trial reports with ≥ 3 errors is unwarranted and not supported by the evidence.

Competing Interests

NF is paid to advise Sanofi, Novo Nordisk and Ipsen on trial methodology and outcomes research.

GR declares non conflicts of interest.

References:

1. Cole GD, Nowbar AN, Mielewczik M, Francis DP. Recognition and Enumeration of Discrepancies of Fact – Leveraged by Any reader to Gauge Security (REDFLAGS): a blinded case-control study. BMJ, in press
2. Nowbar AN, Mielewczik M, Karavassilis M, Dehbi H-M, Shun-Shin MJ, Jones S, et al. Discrepancies in autologous bone marrow stem cell trials and enhancement of ejection fraction (DAMASCENE): weighted regression and meta-analysis. BMJ2014;348:g2688
3. Freemantle N and Rait G. Trials of autologous bone marrow stem cells for heart disease. BMJ 2014;348:g2750
4. Rothman, K.J., Greenland, S., & Lash, T.L. (2008). Modern Epidemiology, 3rd Edition. Philadelphia, PA: Lippincott, Williams & Wilkins.
5. International Conference on Harmonisation of Technical Requirements for Registration of Pharmaceuticals for Human Use. ICH Harmonised Tripartite Guideline: Statistical Principles for Clinical Trials E9
http://www.ich.org/fileadmin/Public_Web_Site/ICH_Products/Guidelines/Efficacy/E9/Step4/E9_Guideline.pdf
6. Anesthesia & Analgesia: <http://journals.lww.com/anesthesia-analgesia/pages/default.aspx>
7. Shafer SL. Shadow of Doubt. Anesthesia & Analgesia 2011; 112: 498-500
8. Seife C. Research Misconduct Identified by the US Food and Drug Administration: Out of Sight, Out of Mind, Out of the Peer-Reviewed Literature. JAMA Intern Med. 2015;175(4):567-577. doi:10.1001/jamainternmed.2014.7774.