

Perhaps it is time to make a paradigm shift and move from calling from the phrase ‘predator journals’ to the more appropriate ‘fake science.’

## 6 UNEXPLAINED DISCREPANCIES BETWEEN PLANNED AND CONDUCTED STATISTICAL ANALYSES IN RANDOMISED TRIALS

<sup>1</sup>Brennan Kahan, <sup>2</sup>Suzie Cro, <sup>3</sup>Gordon Forbes, <sup>2</sup>Nicholas Andrew Johnson. <sup>1</sup>Pragmatic Clinical Trials Unit, Queen Mary University of London, London, UK; <sup>2</sup>Imperial Clinical Trials Unit, Imperial College London, London, UK; <sup>3</sup>Department of Biostatistics and health informatics, Institute of Psychiatry, Psychology and Neuroscience, Kings College London, London, UK

10.1136/bmjebm-2019-EBMLive.6

**Objectives** To evaluate how often pre-specified statistical analysis approaches were publicly available for randomised trials, and the frequency of unexplained discrepancies between planned and conducted analyses.

**Method** We reviewed randomised trials published in six general medical journals (BMJ, JAMA, Lancet, NEJM, PLOS Medicine, and Annals of Internal Medicine) from January-April 2018. Main outcome measures were (i) the number of trials with a publicly available pre-specified analysis approach for the primary outcome (in a protocol or statistical analysis plan [SAP]); (ii) the number of trials with no unexplained discrepancies between the trial publication and the pre-specified approach; and (iii) the types of unexplained discrepancies. Discrepancies were classified as unexplained unless the change was specified in a subsequent version of the protocol or SAP or the discrepancy was discussed in the trial publication. Data extraction was performed independently by two reviewers.

**Results** Overall, 89 of 101 eligible trials (88%) had a publicly available pre-specified analysis approach (83 in a protocol, 6 in a SAP); this document was dated after recruitment began for 27/89 trials (30%), and for 21 trials (24%) no date was available.

Only 22/89 trials (25%) did not have any unexplained discrepancies (n=5 no discrepancies, n=17 explained discrepancies only). Fifty-four trials (61%) had one or more unexplained discrepancies, and in 13 trials (15%) it was impossible to ascertain whether any unexplained discrepancies occurred due to incomplete reporting of the statistical methods in the trial publication. Unexplained discrepancies were most common for the analysis model (n=30, 34%) and analysis population (n=28, 31%), followed by the use of covariates (n=23, 26%) and handling of missing data (n=16, 18%).

Most trials did not report the blinding status of the statistician in relation to database access or final sign-off of the SAP. **Conclusions** Discrepancies in the statistical analysis approach were common. We identified several barriers preventing an evaluation of whether changes may have introduced bias: (i) many protocols and analysis plans were from after recruitment began, preventing a comparison with the pre-trial analysis approach; (ii) discrepancies were rarely explained or justified in the trial publication; (iii) the blinding status of statisticians in relation to modifications of analysis methods was rarely reported; and (iv) some descriptions of the analysis methods used in the final publication were inadequate, preventing a comparison with the pre-specified approach. Resolution of

these barriers is likely to require a multi-faceted approach targeting investigators, journals, and trial registry websites.

## 7 PAPERS AS LIVING DOCUMENTS: USING LITERATE PROGRAMMING TO PRODUCE FULLY TRANSPARENT, REPRODUCIBLE RESEARCH MANUSCRIPTS

<sup>1,2</sup>Matthew Parkes. <sup>1</sup>Arthritis Research UK Centre for Epidemiology, University of Manchester, Manchester, UK; <sup>2</sup>NIHR Manchester Musculoskeletal Biomedical Research Centre, Central Manchester University Hospitals NHS Foundation Trust, Manchester Academic Health Science Centre, Manchester, UK

10.1136/bmjebm-2019-EBMLive.7

**Objectives** The existing model of academic writing and publishing in medical research has not strayed far from its correspondence-based letter writing origins. Authors frequently complain that this system is out of date and restrictive. Currently, articles lack transparency - it is difficult to fully and concisely explain complex analyses without presenting full code and dataset structure. Analyses are also fixed and limited - should readers wish to test assumptions, conduct additional tests, or amalgamate study data with other datasets, they are limited to that which is published, or to contacting authors for datasets and code (often unsuccessfully).

Literate programming is an approach which allows manuscripts to be more transparent, reproducible, and interactive. There are free, open-source tools (e.g. RMarkdown and Pweave) which allow entire research manuscripts to be generated end-to-end in one continuous interwoven block, with live code which runs upon opening the document, which can be compiled into .pdf or HTML.

**Method** This abstract demonstrates how manuscripts can be written using the literate programming model. Using a cleaned, publicly available dataset taken from Vickers' 2006 paper (Vickers *Trials* 2006, 7:15), we use the *knitr*, *rmarkdown*, and *rticles* packages in RStudio, to write a fully transparent example trial results paper. This paper is compiled into .pdf format using the PLOS LaTeX template (*Public Library of Science* 2018; <https://journals.plos.org/plosone/s/latex>), a commonly used manuscript template which is under Creative Commons licence.

**Results** The resulting .pdf manuscript has a format familiar to readers, yet has increased transparency and interactivity, as the code and dataset are distributed with, and are integrated in, the manuscript. The script can be therefore be downloaded and altered to test assumptions, derive values not presented in the paper (to be used in meta-analyses, for example), and reproduce results.

The compiled paper will be available at the author's GitHub repository: <https://github.com/mattyjparkes/EBM-Live2019-example-paper>

**Conclusions** These ‘living papers’, whereby code, data, and interpretation are all interwoven into one live-compiled document has numerous applications for meta-level analyses, and a significantly greater level of transparency. They circumvent the limitations discussed with the current static publishing model, allowing readers to interact and unpick papers in a way that is currently not possible with manuscripts that are divorced from their datasets and analysis code.