

How many researchers use ‘boilerplate’ statistical analysis sections?

An observational study of papers published in *PLOS ONE*.

Nicole White, Richi Nayak, Adrian Barnett

21/05/2020

An ideal statistical analysis will use appropriate methods to create insights from the data and inform the research questions. Unfortunately many current statistical analyses are far from ideal, with many researchers using the wrong methods, misinterpreting the results, or failing to adequately check their assumptions (Goodman 2008; Leek et al. 2017). Some researchers take a “mechanistic” approach to statistics, copying the few methods they know regardless of their appropriateness, and then going through the motions of the analysis (Stark and Saltelli 2018).

Many researchers may not have received adequate training in research methods, and statistics is something they do with trepidation and even ignorance (Altman 1994; King et al. 2019). However, using the wrong statistical methods can cause real harm (Altman 1994) and bad statistical practices are being used abet weak science (Stark and Saltelli 2018). Statistical mistakes are a key source of waste in research and partly explain the current reproducibility crisis in science (Allison et al. 2016). Even when the correct methods are used, many researchers fail to describe them adequately, making it difficult to reproduce the results (Ernst and Albers 2017; Zhou and Skidmore 2018).

The International Committee of Medical Journal Editors recommend that researchers should: “Describe statistical methods with enough detail to enable a knowledgeable reader with access to the original data to judge its appropriateness for the study and to verify the reported results” (ICJME 2019). Although the general lack of statistical understanding from both authors and reviewers means this recommendation may not be checked. A recent survey of editors found that only 23% of health and medical journals used expert statistical review for all articles (Hardwicke and Goodman 2020), which was little different from a survey from 22 years ago (Goodman, Altman, and George 1998).

As statisticians we have heard researchers admit that they sometimes copy-and-paste their statistical methods sections from other papers, regardless of whether they are appropriate. The aim of this paper is to use text-mining methods to estimate the extent that researchers are using ‘boilerplate’ statistical methods sections. Use of these methods sections indicates that little thought has gone into the statistical analysis.

Methods

Data sources

Public Library of Science (PLOS ONE)

PLOS ONE is a large open access journal that publishes original research across the fundamental sciences. Article submissions are handled by an academic editor who selects peer reviewers based on their self-nominated area(s) of expertise. Submissions do not undergo formal statistical review. Instead, reviewers are required to assess submissions against several publication criteria, including whether “Experiments, statistics, and other analyses are performed to a high technical standard and are described in sufficient detail”. Authors are encouraged to follow published reporting guidelines to ensure that chosen statistical methods are appropriate

for the study design, and adequate details are provided to enable independent replication of results. For example, the EQUATOR network has developed reporting guidelines for study designs commonly used in health research, in part because of the long history of poor statistical application and reporting in health and medical journals (Altman and Simera 2016).

All articles published in PLOS ONE are freely accessible via the PLOS Application Programming Interface (API). This functionality enabled us to conduct semi-automated searches of full-text articles and analyse data on individual records, including text content and general attributes such as publication date and field(s) of research. Our search strategy comprised of targeted API searches followed by article filtering based on available section headings.

Step 1: Targeted API searches. API searches were completed using the R package ‘rplos’ (Chamberlain, Boettiger, and Ram 2020). Search queries targeted the presence of analysis-related terms anywhere within a full-text article. Individual search terms combined the words “data” or “statistical”, with one of: “analysis”, “analyses”, “method”, “methodology” or “model(l)ing”. Search terms were intended to be broad whilst keeping search results to a manageable number for full-text review (see Step 2). By allowing terms to appear anywhere within the article, we accounted for the possibility of relevant text being placed in different sections, for example, in the Material and Methods section versus Supplementary Information. Search results were indexed by a unique Digital Object Identifier (DOI). Attribute data collected per DOI included journal volume, subject classification(s) and total article views since publication.

Step 2: Partial matching on section headings. Full text XML data for all search results were downloaded and combined into a single dataset, organised by DOI and subsection heading(s). Since PLOS ONE does not prescribe standardised headings to preface statistical methods sections, we performed partial matching on available headings against frequently used terms in initial search results: ‘Statistical analysis’, ‘Statistical analyses’, ‘Statistical method’, ‘Statistics’, ‘Data analysis’ and ‘Data analyses’. To determine the reliability of our chosen filters, we manually reviewed full text data extracted for a random sample of XXX articles that were not matched (File S1).[TODO...finish this thought...]

Australia and New Zealand Clinical Trials Registry (ANZCTR)

ANZCTR was established in 2005 as part of a coordinated global effort to improve research quality and transparency in clinical trials reporting. Researchers responsible for overseeing a clinical trial are expected to register full details before commencing participant enrolment. Details required for registration follow a standardised template (reference or supp file), which covers important details about participant eligibility, intervention(s) being evaluated, study design and outcomes. As part of study design, researchers are required to provide a summary of statistical methods and planned analyses. Upon successful registration, the trial is assigned a unique registration number for use in future reporting in scientific publications. All trials registered on ANZCTR are publicly available and can be searched for via an online portal. [TODO: details of how stat.section text was extracted for analysis]

Full-text processing

Text cleaning aimed to standardise notation and statistical terminology, whilst minimising changes to article style and formatting. Full details of text cleaning steps undertaken and corresponding R code are provided in Supplementary File X.

Mathematical notation including Greek letters was converted from Unicode characters to plain text. For example, the Unicode characters corresponding to θ (<U+03B8>) was replaced with ‘theta’. Similarly, common symbols outside of Unicode blocks including ‘%’ (percent) and ‘<’ (‘less-than’) were converted into plain text, using functions available in the ‘textclean’ package (Rinker 2018). General formatting including carriage returns, punctuation marks, in-text references (e.g. [42]) centred equations and other non-ascii characters were removed. Text contained inside brackets was retained in the dataset to maximise content for analysis, with brackets removed.

We compiled an extensive list of statistical terms to standardise descriptions of statistical methods reported across included record in both datasets. An initial list was compiled by calculating individual word frequencies and identifying relevant terms that appeared at least 100 times. Further terms were sourced from index searches of statistics reference textbooks [ref]. The final list is provided as Supplementary Material. Possible variants including plurals (e.g. ‘chi-squares’) unhyphenated (e.g chi square) and combined (e.g. chisquare) terms were transformed to singular, hyphenated form (e.g chi-square). Common statistical tests were also hyphenated (e.g. hosmer lemeshow to hosed-lemeshow).

As a final step, common stop words including pronouns, contractions and selected prepositions were removed. We chose to keep certain stop words that, if excluded, may have changed the context of statistical method(s) being described (e.g. ‘between’, ‘against’).

Clustering algorithm

Results

PLOS ONE

- Final sample size
- Breakdown of included articles by year of publication (did/did not include stats section). Final sample broken down by subject classification(s) [most are biology and life sciences]
- Figure showing change in gof measure as function of number of clusters (use to justify choice of number of clusters)
- Describe topic clusters + examples

ANZCTR

- Final sample size

Discussion

The first line in many statistical analysis sections was the software used, implying that the software is the most important detail. As Doug Altman said, “Many people think that all you need to do statistics is a computer and appropriate software” (Altman 1994). This is not the case, and whilst it is important for researchers to mention the software and version used for reproducibility purposes, it is a relatively minor detail compared with detailing what methods were used and why.

Despite the extensive array of tests available, many authors are reporting the same few methods.

Limitations

We did not check whether papers used the correct methods, and for some simple studies a ‘boilerplate’ statistical methods section would be fine.

We examined papers where there was a statistics section, and we missed papers that used statistical analysis but did not include a statistical analysis section. Reiterate outcomes of random sample checking here.

We only examined one journal and hence our results may not be generalisable to all journals, especially those that use a statistical reviewer for all papers.

References

- Allison, David B., Andrew W. Brown, Brandon J. George, and Kathryn A. Kaiser. 2016. "Reproducibility: A Tragedy of Errors." *Nature* 530 (7588): 27–29. <https://doi.org/10.1038/530027a>.
- Altman, D G. 1994. "The Scandal of Poor Medical Research." *BMJ* 308 (6924): 283–84. <https://doi.org/10.1136/bmj.308.6924.283>.
- Altman, Douglas G, and Iveta Simera. 2016. "A History of the Evolution of Guidelines for Reporting Medical Research: The Long Road to the EQUATOR Network." *Journal of the Royal Society of Medicine* 109 (2): 67–77. <https://doi.org/10.1177/0141076815625599>.
- Chamberlain, Scott, Carl Boettiger, and Karthik Ram. 2020. *Rplos: Interface to the Search Api for 'Plos' Journals*. <https://CRAN.R-project.org/package=rplos>.
- Ernst, Anja F., and Casper J. Albers. 2017. "Regression Assumptions in Clinical Psychology Research Practicea Systematic Review of Common Misconceptions." *PeerJ* 5: e3323. <https://doi.org/10.7717/peerj.3323>.
- Goodman, Steven. 2008. "A Dirty Dozen: Twelve P-Value Misconceptions." *Seminars in Hematology* 45 (3): 135–40. <https://doi.org/10.1053/j.seminhematol.2008.04.003>.
- Goodman, Steven N., Douglas G. Altman, and Stephen L. George. 1998. "Statistical Reviewing Policies of Medical Journals." *Journal of General Internal Medicine* 13 (11): 753–56. <https://doi.org/10.1046/j.1525-1497.1998.00227.x>.
- Hardwicke, Tom E, and Steve Goodman. 2020. "How Often Do Leading Biomedical Journals use Statistical Experts to Evaluate Statistical Methods? The Results of a Survey." *MetaArXiv*. <https://doi.org/10.31222/osf.io/z27u4>.
- ICJME. 2019. "Recommendations for the Conduct, Reporting, Editing, and Publication of Scholarly Work in Medical Journals." <http://www.icmje.org/icmje-recommendations.pdf>.
- King, Kevin M., Michael D. Pullmann, Aaron R. Lyon, Shannon Dorsey, and Cara C. Lewis. 2019. "Using Implementation Science to Close the Gap Between the Optimal and Typical Practice of Quantitative Methods in Clinical Science." *Journal of Abnormal Psychology* 128 (6): 547–62. <https://doi.org/10.1037/abn0000417>.
- Leek, Jeff, Blakeley B. McShane, Andrew Gelman, David Colquhoun, Michèle B. Nuijten, and Steven N. Goodman. 2017. "Five Ways to Fix Statistics." *Nature* 551 (7682): 557–59. <https://doi.org/10.1038/d41586-017-07522-z>.
- Rinker, Tyler W. 2018. *textclean: Text Cleaning Tools*. Buffalo, New York. <https://github.com/trinker/textclean>.
- Stark, Philip B., and Andrea Saltelli. 2018. "Cargo-Cult Statistics and Scientific Crisis." *Significance* 15 (4): 40–43. <https://doi.org/10.1111/j.1740-9713.2018.01174.x>.
- Zhou, Yuanyuan, and Susan Skidmore. 2018. "A Reassessment of Anova Reporting Practices: A Review of Three Apa Journals." *Journal of Methods and Measurement in the Social Sciences* 8 (1): 3–19. <https://doi.org/10.2458/v8i1.22019>.