RESEARCH ARTICLE

OPEN ⦵ ACCESS

A snapshot of statistical methods used in experimental immunoblotting: a scoping review

Romain-Daniel Gosselin*

Precision Medicine Unit, Lausanne University Hospital, Chemin des Roches 1a/1b, 1010 Lausanne, Switzerland

Abstract − *Background*: Among the many avenues considered to make life science more reproducible, the improvement of the quality and openness of statistical methods has taken centre stage. However, although disparities across research fields and techniques are not unknown, they remain largely unexplored. *Objectives*: Provide an overview of statistical methods used in biochemical research involving immunoblotting (also referred to as western blotting), a technique frequently used to quantify proteins of interest. *Source of evidence*: PubMed. *Eligibility criteria*: Studies reporting immunoblots with quantitative interpretation (statistical inference). *Charting Methods*: A reverse chronological systematic sampling was implemented to analyse 2932 experimental conditions (i.e., experimental groups) from 64 articles published at the end of 2021. The statistical test (actual study size $n = 67$) and software (actual study size $n = 61$) used for each article and the sample size for each experimental condition were documented. *Results*: The results indicate an overhelming number of parametric tests, mostly one-way analysis of variance (ANOVA, 15/67) and Student's *t*-test (13/67), but for many articles the statistical procedure was not clearly stated (23/67). GraphPad Prism was the most commonly used statistical package (36/61), but many (14/61) articles did not reveal the package used. Finally, the sample size was disclosed in only 1054/2932 conditions in which its median value was 3 (IQR = [3–6]). *Conclusion*: This study suggests that the transparency of reporting might be suboptimal in immunoblotting research and prompts the need for more comprehensive reviews in the future.

Keywords: Biochemistry, Good practices, Meta-research, Publishing, Reporting, Reproducibility, Statistics, Western blot

Introduction

Transparency, correctness, and quality in published protocols are key prerequisites to ensure replication in science and to enable evidence-based research practice. In particular, the use of inappropriate statistical designs and methods as well as their insufficient reporting are considered important contributors to the so-called reproducibility crisis in life science [1]. Guidelines describing statistical reporting are regularly released, thereby establishing an important framework of standardization for published material. These include Animal Research: Reporting of In Vivo Experiments (ARRIVE, recently updated as ARRIVE 2.0) [2, 3], the Checklist for Reporting In-vitro Studies (CRIS) [4], guidelines from the American Physiological Society [5] and the checklist by *Emmerich and Harris* for *in vitro* research [6]. Although the aforementioned guidelines are generally addressed to a well-defined scientific community, the current statistical practices in reporting across different research domains remain largely unknown. In this context, meta-research that describes the characteristics of statistical methods in specific research fields would ultimately enable the implementation of targeted educational initiatives in defined research communities.

Immunoblotting (also referred to as "western blotting") is a very common technique in biochemistry that uses protein separation by molecular weight using Sodium dodecyl-sulfate polyacrylamide gel electrophoresis (SDS-PAGE) and a subsequent antibody-based specific detection and relative quantification of proteins of interest. Others have explained that method disclosure in research based on antibodies, including immunoblotting, was not sufficient to enable reproducibility, which prompted the release of specific reporting guidelines in the field [7, 8]. However, these recommendations focused primarily on the disclosure of reagents, laboratory protocols and methods of image acquisition/processing with only minor sections on statistical analysis of quantitative blots.

The present study aims to assess statistical methods used in immunoblotting to rationalize further actions such as comprehensive systematic reviews or the assembling of specific guidelines.

*Corresponding author: Romain-Daniel.Gosselin@chuv.ch

Material and methods

This study was designed and prepared according to the 2018 Preferred Reporting Items for Systematic reviews and Meta-Analyses extension for Scoping Reviews (PRISMA-ScR) [9] and the 2020 "*Updated methodological guidance for the conduct of scoping reviews*" [10, 11]. The dual indication was to "examine how research is conducted on a certain topic or field" and serve "as a precursor to a systematic review". In addition, this manuscript was also prepared in accordance with the ACcess to Transparent Statistics (ACTS) call to action [12]. Therefore, the methods section starts with the statistical paragraph that includes the eight components of ACTS and the discussion incorporates a specific section dedicated to the limitations of the study. For the sake of clarity, throughout the text, the term "sample size" refers to the number of experimental conditions in immunoblots in the articles included in the review, and the term "study size" refers to the number of analytical units in the present review (either the number of articles or total number of experimental conditions included). The study protocol was not preregistered.

Statistical methods

Data were collected, structured, and managed using Microsoft Excel for Mac (version 16.59). Calculations of descriptive statistics (quartiles, interquartile range) and the drawing of bar charts and histograms were performed using the base function in R software (R-studio, version 4.1.1). The statistical items quantified in the study were: the software used, the procedures (tests) implemented, and the number of observations collected (i.e., sample size). The analytical units for the documentation of software and tests were the articles (i.e., the items were listed for each article). For the assessment of sample size, the analytical units were the separate observations (bands), presumed independent, used by the authors in each condition. The study size was determined a priori using a precision-based approach by applying the following formula:

$$N = \frac{Z^2 p(1-p)}{d^2},$$

N is the minimum number of articles required to establish a 95% confidence interval (CI, $Z = 0.95$) for software and tests if the least frequent software or test has a frequency (p) of 0.1 (i.e., present in 10% of articles), with a precision (d) of 0.07 (i.e., 7%). The calculation gave a minimum size of 70 articles for the study.

Article sampling

Articles with quantitative western blots were sampled from the PubMed browser of the NCBI website (https://pubmed.ncbi.nlm.nih.gov/) using reverse chronological convenience sampling (Fig. 1, eligibility criteria are presented in Tab. 1). In the advanced NCBI research tool, the following terms were entered for query: "immunoblot" OR "western blot" OR "western-blot" OR "sds page" OR

"sds-page" and the NCBI publication filter was used with "end date" set at 31.12.2021. Retrieved articles were individually checked for compliance with inclusion criteria and absence of exclusion criteria by clicking on the link to the journal page or DOI until 70 articles were collected. In particular, the article text and figures were first examined to confirm both that the authors performed an immunoblot (and not a mere polyacrylamide gel electrophoresis) and that at least one reported blot was quantitative (presence of statistics and/or of text providing a quantitative interpretation of the blots). Twenty-six publications were excluded at this stage. During the subsequent detailed analysis of statistical features (see infra), six additional articles were excluded because they did not perform quantitative immunoblots, contrary to what was presumed upon first reading.

Quantification of statistical features

For each article sampled and for the experiments that reported quantitative immunoblots, the corresponding text (methods, results, discussion, figure legends) was explored, and three statistical items were documented: the statistical software used, the statistical procedure (e.g., test) utilized and the sample size (disclosed number of observations) in each experimental condition. Immunoblots in figures provided as supplemental materials were not included, but supplemental methods referring to figures in the main text were considered full-fledged methodological information. The category "not disclosed/unclear" corresponds to instances where the item was not given, although a statistical analysis (e.g., p-value, discussion on statistical significance) was provided, or was not given in an unambiguous manner. Absence of clear information about a post-test used after ANOVA was included in this latter category.

Data and code availability

All data generated and analysed during this study, as well as the R code used, are publicly available on Figshare at https://doi.org/10.6084/m9.figshare.19448264.v1.

Results

The final list of included ($n = 64$) and excluded articles is provided on the Figshare public repository (https://doi.org/10.6084/m9.figshare.19448264.v1).

Out of the 64 sampled articles, two articles used multiple procedures to analyse their immunoblots. One publication used both Student's t-test and an unknown post-test after an ANOVA, and no statistical analysis for a blot was interpreted quantitatively. Another article used both an unclear test for one immunoblot figure and no statistical analysis for a western blot interpreted quantitatively. Therefore, the final study size used in the present description of statistical tests used is $n = 67$. On the other hand, three sampled articles did not provide any statistical analysis or statistical software although they reported quantitative analyses, therefore the final study size for the documentation of software is $n = 61$ (Fig. 1).
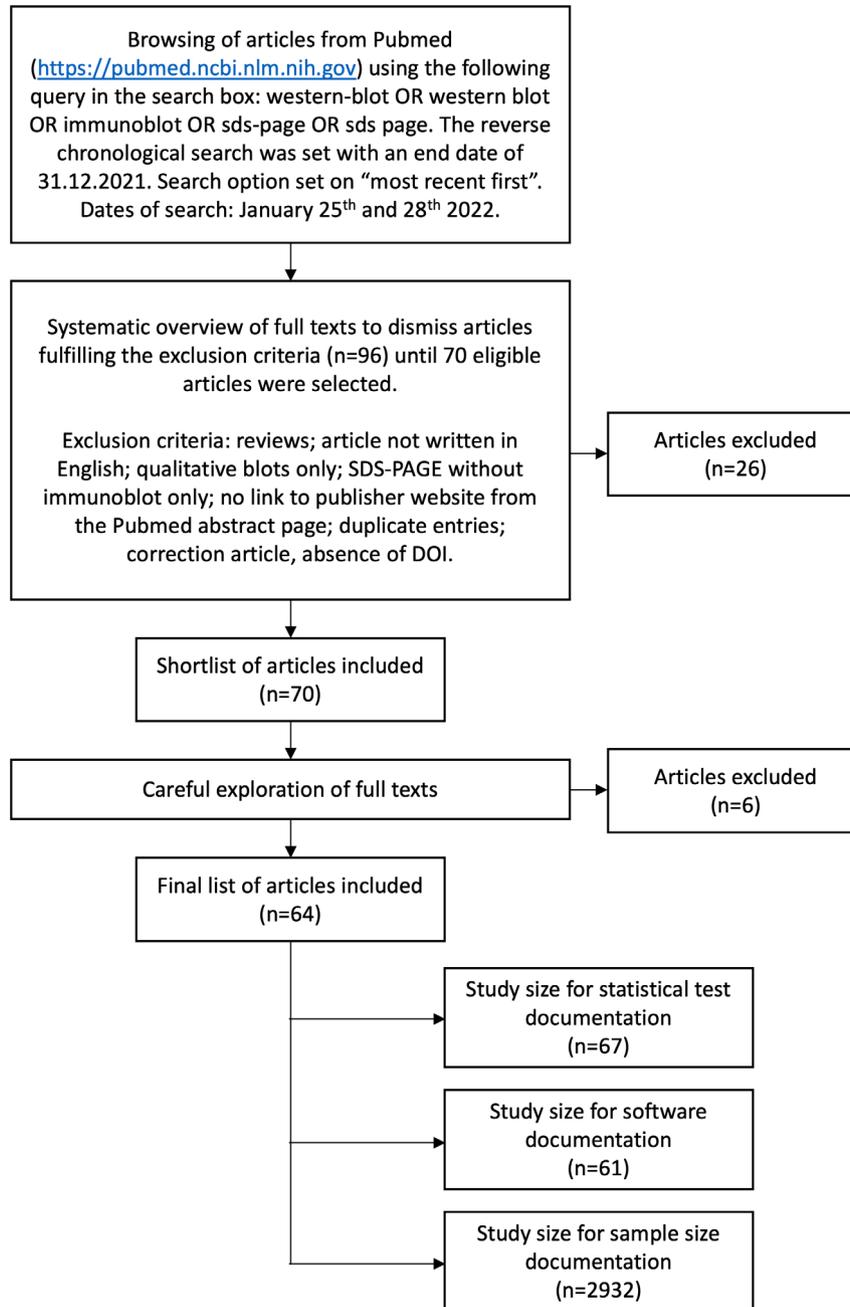
Browsing of articles from Pubmed (https://pubmed.ncbi.nlm.nih.gov) using the following query in the search box: western-blot OR western blot OR immunoblot OR sds-page OR sds page. The reverse chronological search was set with an end date of 31.12.2021. Search option set on "most recent first". Dates of search: January 25th and 28th 2022.

Systematic overview of full texts to dismiss articles fulfilling the exclusion criteria (n=96) until 70 eligible articles were selected.

Exclusion criteria: reviews; article not written in English; qualitative blots only; SDS-PAGE without immunoblot only; no link to publisher website from the Pubmed abstract page; duplicate entries; correction article, absence of DOI.

Articles excluded (n=26)

Shortlist of articles included (n=70)

Careful exploration of full texts

Articles excluded (n=6)

Final list of articles included (n=64)

Study size for statistical test documentation (n=67)

Study size for software documentation (n=61)

Study size for sample size documentation (n=2932)

Figure 1. Flow chart of the sampling protocol.

Table 1. Eligibility criteria applied to the PubMed search engine.

End date	31.12.2021
Boolean search	Western-blot OR western blot OR immunoblot OR sds page OR sds-page
Inclusion criteria	Presence of quantitative western blot (i.e., graph, table or text presenting a quantification)
Exclusion criteria	Article not written in English; qualitative western blots only (e.g., expression control, specificity...); SDS-PAGE without immunoblotting; DOI or link to publisher's website not displayed on PubMed; review article.

The descriptive analysis of statistical procedures is presented in Figure 2a. The results revealed widespread use of parametric tests, namely one-way analysis of variance (ANOVA) with post-test (15/67), Student's t-test (regardless of their laterality or paired vs. unpaired nature, 13/67) and two-way ANOVA with post-test (5/67).

A

Statistical tests
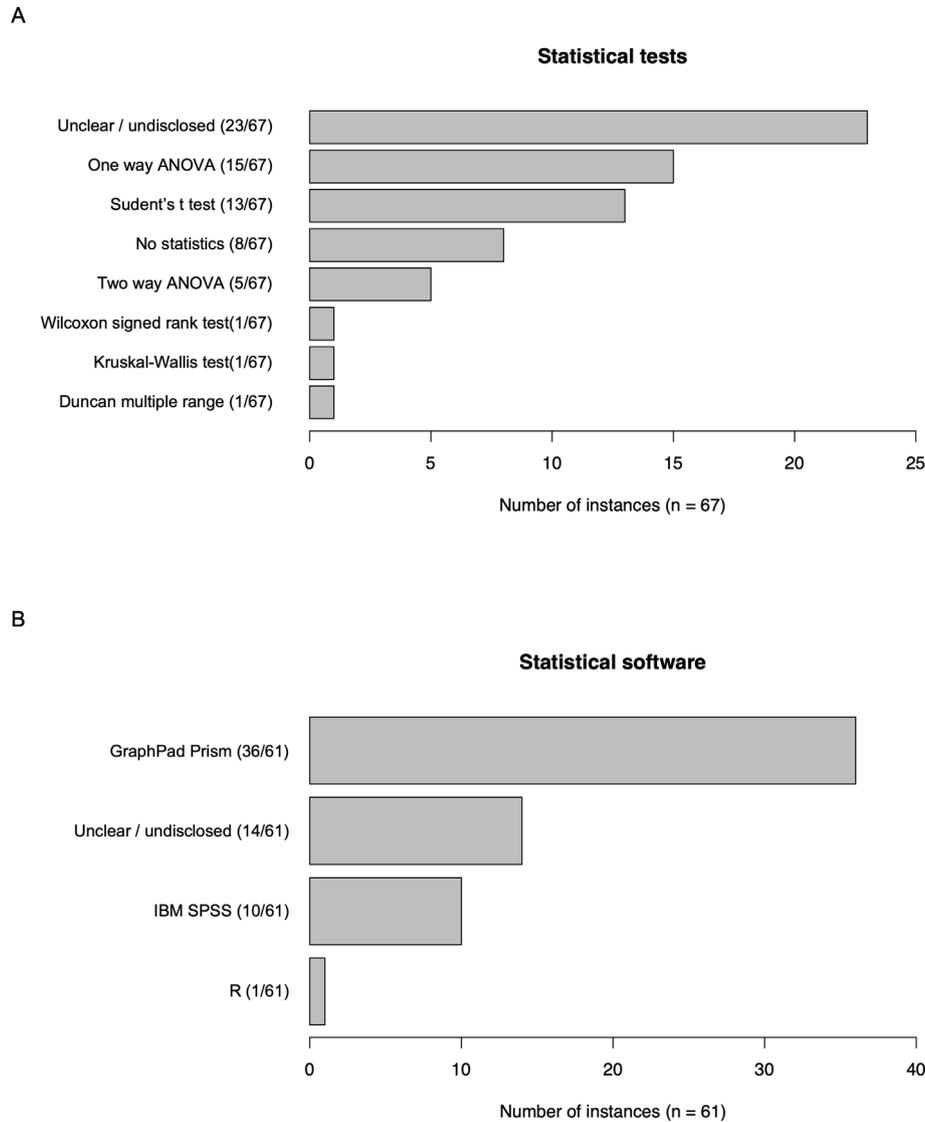

B

Statistical software

Figure 2. Quantification of statistical methods (a, study size: $n = 67$) and software (b, study size: $n = 61$) in articles ($n = 64$) with quantitative immunoblotting.

Non-parametric tests were mostly absent, with only Wilcoxon signed-rank test ($1/67$), Kruskal–Wallis test (followed by a post-test $1/67$), and Duncan's multiple range test ($1/67$). Strikingly, in a large proportion of quantitative analyses, the statistical procedure was unclear or not disclosed ($23/67$, which included instances where the post-test was not disclosed after the omnibus procedure), or no statistical analyses were performed ($8/67$). The evaluation of statistical software shows that only three packages were cited (Fig. 2b). The most frequently used was GraphPad Prism ($36/61$), followed by IBM SPSS ($10/61$) and the open-source software R ($1/61$). Fourteen articles ($14/61$) did not cite any software.

The sample size was clearly indicated in only $1054/2932$ experimental conditions and was conversely often unclear (e.g., when inequalities were given instead of exact numbers) or not disclosed ($1592/2932$). The frequency distribution of sample sizes (Fig. 3) shows the vast over-representation of very small samples, with $744/1054$ experimental conditions that incorporated fewer than four observations per sample. In about 10% of experimental conditions ($286/2932$), the notion of sample size was even irrelevant due to the use of single blots to quantitatively interpret the experiment. The analysis of disclosed sample sizes shows that the median number of observations (i.e., individual blots) used in quantitative immunoblots is 3 (range [2–40], interquartile range, IQR [3–6]).

Discussion

In the sample analysed in the present scoping review, the quantitative immunoblotting was typically designed with small numbers of observations and analysed with parametric tests using proprietary software. The use of such designs jeopardize their validity and sensitivity
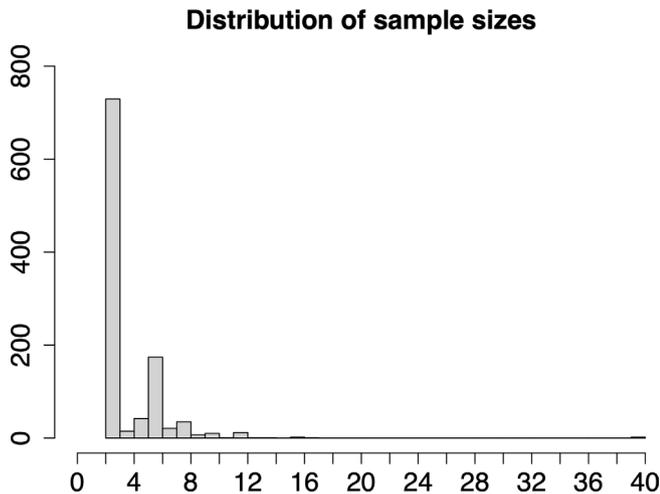
Distribution of sample sizes

Figure 3. Histogram showing the frequency (Y axis) distributions of sample size (X axis) in immunoblotting experimental conditions (actual study size: $n = 1054$). Only experimental conditions with clearly disclosed sample sizes were included (1054/2932).

(i.e., statistical power). First, the verification of parametric assumptions, which are a mandatory prerequisite for the use of parametric tests, is challenging when samples are small. Indeed, in that context, graphical methods such as quantile–quantile (Q–Q) plots are also not reliable and distribution tests for normality (e.g., Shapiro–Wilk, Kolmogorov–Smirnov) tend to be largely underpowered to detect any deviation from normality. Furthermore, statistically significant differences obtained with small samples tend to overestimate effect sizes and to have a higher percentage of false positives among statistically significant results [13]. In addition, small sample sizes such as observed herein tend also to have low statistical power and thus inflate the fraction of false negative results. This shrinkage of statistical power may even engender an impossibility to detect statistically significant differences at the usual confidence levels if non-parametric tests are used [14]. Therefore, for all these reasons, the sampled literature that reports quantitative immunoblotting might be riddled with a large share of statistically unreliable results.

In the present sample, a large proportion of statistical methods were not disclosed, data that is in line with the lack of transparency recurrently reported in other studies [15–17], which is a strong impediment to reproducibility. Many guidelines exist but their publication and dissemination has shown limited impact. Therefore, although this route of action appears necessary, especially to provide a framework of official recommendations, it is probably not sufficient to initiate a change towards a more reproducible culture. Strengthening continuing education programmes on good research practices for graduate students, postdoctoral researchers and established researchers would be an important avenue to explore. The organisation of such education for professionals at career stages where improving methodological and statistical literacy relates to the actual daily laboratory activity might be more efficient than

targeting young students at the undergraduate level. Furthermore, in methodological courses, a marked emphasis should be placed on experimental design and data sampling rather than mere data analysis and statistical inference [18].

The question of whether research involving immunoblotting uses different standards of design, analytic methods and reporting than other fields of experimental science would be unequivocally answered by more comprehensive systematic reviews. Should it be the case, measures assembled to particularly target this research community might be warranted, including guidelines in field-specific journals, thematic sessions on design and transparency in scientific events or more active education on these topics in biochemistry curricula.

Importantly, the alarmingly high prevalence of non-transparent methods points also to a similar insufficient statistical literacy among peer reviewers in the field. Although guidelines articles intended for peer reviewers are available [7, 19], peer reviewing and editorial filters do not apparently constitute efficient bottlenecks for statistical quality. One solution might be the systematic invitation of statistical reviewers to help ensure standards in statistical reporting while, at the same time, lifting some of the reviewing burden from non-statistical reviewers [20]. Nevertheless, it is conceivable to restrict this option to manuscripts with relatively complex designs, leaving non-statistical reviewers to perform an assessment of statistical transparency and elementary statistics in most studies. In this respect, standardization of minimal statistical knowledge among reviewers is important and would benefit from the widespread creation of continuing education programmes in manuscript peer reviewing [21]. Authoring and reviewing activities are fulfilled by overlapping groups of scholars, and the concepts to be assimilated are the same. Therefore, it is also suggested here that guidelines on reporting practices should systematically be clearly addressed to both researchers and peer reviewers and make explicit that better knowledge of such sound practices is important when engaging in either activity. In addition, specific educational programmes on biostatistics and reporting organized in institutes that perform immunoblotting research should similarly unambiguously clarify that the exposed concepts apply similarly to the design, analysis, and presentation of a research project and its assessment as a reviewer.

Limitations

The present review has several limitations. First, other statistical features could have been considered to achieve a more comprehensive description of statistical reporting. These include the choice of error, the nature of post-tests, or the significance thresholds used. In addition, this study was designed as descriptive and does not intend to make inferences about all the literature that uses western blots. For example, the date of publication selected and the convenience sampling method may have created an over-representation of some journals and editors that prevents

generalization. Nevertheless, the consistency of the results within this sample is already notable, suggesting a widespread entrenched culture of incomplete reporting in immunoblot research and prompts for future confirmatory systematic review. Since not all periodicals that publish immunoblots are referenced in PubMed, the use of this sole database may also be a source of bias. Furthermore, the study protocol was not preregistered, which constitutes another limitation. Finally, the present study was designed and analysed by a single reviewer, a methodology deemed acceptable for accelerated evidence synthesis such as the current review [22, 23] but which might increase error and bias for comprehensive systematic reviews [23, 24]. Therefore, future systematic reviews would ideally involve two reviewers, use a large sample of articles from at least two databases and a sampling scheme that enable the random collection of immunoblot studies on a larger time period (e.g., 6 months) and from many distinct journals.

Conclusion

In conclusion, the results indicate that small samples, parametric tests and proprietary software might be omnipresent in immunoblotting research and that non-transparency in statistical protocols is widespread. Targeting researchers in this specific field with continuing education on good research practice and fundamental biostatistics should be a priority objective. Additional avenues to explore are better communication or enforcement of existing reporting guidelines. Journals and editors should take a leading role in these efforts.

Conflict of interest

Acknowledgments

Author contributions

RDG designed the study, analysed the data, and wrote the manuscript.

References

1. Prager EM, Chambers KE, Plotkin JL, McArthur DL, Bandrowski AE, Bansal N, Martone ME, Bergstrom HC, Bespalov A, Graf C (2019), Improving transparency and scientific rigor in academic publishing. J Neuro Res 97, 377–390. https://doi.org/10.1002/jnr.24340.

2. Kilkenny C, Browne WJ, Cuthill IC, Emerson M, Altman DG (2010), Improving bioscience research reporting: the ARRIVE guidelines for reporting animal research. PLoS Biol 8, e1000412. https://doi.org/10.1371/journal.pbio.1000412.

3. Percie du Sert N, Hurst V, Ahluwalia A, Alam S, Avey MT, Baker M, Browne WJ, Clark A, Cuthill IC, Dirnagl U, Emerson M, Garner P, Holgate ST, Howells DW, Karp NA, Lazic SE, Lidster K, MacCallum CJ, Macleod M, Pearl EJ, Petersen OH, Rawle F, Reynolds P, Rooney K, Sena ES, Silberberg SD, Steckler T, Wurbel H (2020), The ARRIVE guidelines 2.0: Updated guidelines for reporting animal research. PLoS Biol 18, e3000410. https://doi.org/10.1371/journal.pbio.3000410.

4. Krithikadatta J, Gopikrishna V, Datta M (2014), CRIS Guidelines (Checklist for Reporting In-vitro Studies): A concept note on the need for standardized guidelines for improving quality and transparency in reporting in-vitro studies in experimental dental research. J Conserv Dent 17, 301–304. https://doi.org/10.4103/0972-0707.136338.

5. Yosten GLC, Adams JC, Bennett CN, Bunnett NW, Scheman R, Sigmund CD, Yates BJ, Zucker IH, Samson WK (2018), Revised guidelines to enhance the rigor and reproducibility of research published in American Physiological Society journals. Am J Physiol Regul Integr Comp Physiol 315, R1251–R1253. https://doi.org/10.1152/ajpregu.00274.2018.

6. Emmerich CH, Harris CM (2019), Minimum information and quality standards for conducting, reporting, and organizing in vitro research. Handbook of Experimental Pharmacology 257, 177–196. https://doi.org/10.1007/164_2019_284.

7. Brooks HL, Lindsey ML (2018), Guidelines for authors and reviewers on antibody use in physiology studies. Am J Physiol Heart Circ Physiol 314, H724–H732. https://doi.org/10.1152/ajpheart.00512.2017.

8. Gilda JE, Ghosh R, Cheah JX, West TM, Bodine SC, Gomes AV (2015), Western Blotting inaccuracies with unverified antibodies: need for a Western Blotting Minimal Reporting Standard (WBMRS). PLoS ONE 10, e0135392. https://doi.org/10.1371/journal.pone.0135392.

9. Tricco AC, Lillie E, Zarin W, O'Brien KK, Colquhoun H, Levac D, Moher D, Peters MDJ, Horsley T, Weeks L, Hempel S, Akl EA, Chang C, McGowan J, Stewart L, Hartling L, Aldcroft A, Wilson MG, Garritty C, Lewin S, Godfrey CM, Macdonald MT, Langlois EV, Soares-Weiser K, Moriarty J, Clifford T, Tuncalp O, Straus SE (2018), PRISMA Extension for Scoping Reviews (PRISMA-ScR): checklist and explanation. Ann Intern Med 169, 467–473. https://doi.org/10.7326/M18-0850.

10. Munn Z, Peters MDJ, Stern C, Tufanaru C, McArthur A, Aromataris E (2018), Systematic review or scoping review? Guidance for authors when choosing between a systematic or scoping review approach. BMC Med Res Methodol 18, 143. https://doi.org/10.1186/s12874-018-0611-x.

11. Peters MDJ, Marnie C, Tricco AC, Pollock D, Munn Z, Alexander L, McInerney P, Godfrey CM, Khalil H (2020), Updated methodological guidance for the conduct of scoping reviews. JBI Evid Synth 18, 2119–2126. https://doi.org/10.11124/JBIES-20-00167.

12. Gosselin RD (2020), Statistical analysis must improve to address the reproducibility crisis: The ACcess to Transparent Statistics (ACTS) call to action. Bioessays 42, e1900189. https://doi.org/10.1002/bies.201900189.

13. Pawitan Y, Michiels S, Koscielny S, Gusnanto A, Ploner A (2005), False discovery rate, sensitivity and sample size for microarray studies. Bioinformatics 21, 3017–3024. https://doi.org/10.1093/bioinformatics/bti448.

14. Krzywinski M, Altman N (2014), Points of significance: Nonparametric tests. Nat Methods 11, 467–468. https://doi.org/10.1038/nmeth.2937.

15. Avey MT, Moher D, Sullivan KJ, Fergusson D, Griffin G, Grimshaw JM, Hutton B, Lalu MM, Macleod M, Marshall J, Mei SH, Rudnicki M, Stewart DJ, Turgeon AF, McIntyre L, Canadian Critical Care Translational Biology G (2016), The devil is in the details: incomplete reporting in preclinical animal research. PLoS One 11, e0166733. https://doi.org/10.1371/journal.pone.0166733.

16. Gosselin RD (2021), Insufficient transparency of statistical reporting in preclinical research: a scoping review. Nature 11, 3335. https://doi.org/10.1038/s41598-021-83006-5.

17. Weissgerber TL, Garcia-Valencia O, Garovic VD, Milic NM, Winham SJ (2018), Why we need to report more than "Data were Analyzed by t-tests or ANOVA". Elife 7, e36163, https://doi.org/10.7554/eLife.36163.

18. Reynolds PS (2022), Between two stools: preclinical research, reproducibility, and statistical design of experiments. BMC Res. Notes 15, 73. https://doi.org/10.1186/s13104-022-05965-w.

19. Morton JP (2009), Reviewing scientific manuscripts: how much statistical knowledge should a reviewer really know? Adv Physiol Educ 33, 7–9. https://doi.org/10.1152/advan.90207.2008.

20. Cobo E, Selva-O'Callagham A, Ribera JM, Cardellach F, Dominguez R, Vilardell M (2007), Statistical reviewers improve reporting in biomedical articles: a randomized trial. PLoS One 2, e332. https://doi.org/10.1371/journal.pone.0000332.

21. Kawczak S, Mustafa S (2020), Manuscript review continuing medical education: a retrospective investigation of the learning outcomes from this peer reviewer benefit. BMJ Open 10, e039687. https://doi.org/10.1136/bmjopen-2020-039687.

22. Gartlehner G, Affengruber L, Titscher V, Noel-Storr A, Dooley G, Ballarini N, Konig F (2020), Single-reviewer abstract screening missed 13 percent of relevant studies: a crowd-based, randomized controlled trial. J Clin Epidemiol 121, 20–28. https://doi.org/10.1016/j.jclinepi.2020.01.005.

23. Waffenschmidt S, Knelangen M, Sieben W, Buhn S, Pieper D (2019), Single screening versus conventional double screening for study selection in systematic reviews: a methodological systematic review. BMC Med Res Methodol 19, 132. https://doi.org/10.1186/s12874-019-0782-0.

24. Stoll CRT, Izadi S, Fowler S, Green P, Suls J, Colditz GA (2019), The value of a second reviewer for study selection in systematic reviews. Res Synth Methods 10, 539–545. https://doi.org/10.1002/jrsm.1369.