EXCLI Journal 2026;25:1583-1587 – ISSN 1611-2156
Received: September 30, 2026, accepted: October 02, 2026, published: October 21, 2026

## Letter to the editor:

# IMPROVING TRANSLATIONAL ONCOLOGY WITH BETTER STATISTICS

Siti Nazihahasma Hassan[1,2*], Wan Muhamad Amir W Ahmad[3*]

[1] Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia
[2] Hospital Pakar Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia
[3] Biostatistics Unit, School of Dental Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia

* **Corresponding authors:** Siti Nazihahasma Hassan, Department of Neurosciences, School of Medical Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia; E-mail: sitinazihah20@usm.my
Wan Muhamad Amir W Ahmad, Biostatistics Unit, School of Dental Sciences, Universiti Sains Malaysia, 16150 Kubang Kerian, Kelantan, Malaysia; E-mail: wmamir@usm.my

https://dx.doi.org/10.17179/excli2026-9843

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (https://creativecommons.org/licenses/by/4.0/).

## ABSTRACT

In most empirical studies, each research question is assessed using statistical hypothesis testing, with a *P* value below the pre-specified significance level (α) providing evidence against the null hypothesis. Whilst reporting *P* values is common practice, they may only indicate the presence or absence of evidence for a difference. Of the 129 preclinical anticancer studies published in 2025 that were reviewed, 95 performed null hypothesis significance testing, including analysis of variance and/or *t* tests. Of these, only two reported effect size measures. In particular, several aspects of statistical reporting could be improved in future research to facilitate evidence synthesis and translational oncology.

**Keywords:** *P* value, anticancer research, statistical analysis, statistical reporting, biomedical studies

The quality of statistical reporting in biomedical journals remains inadequate. Several recurring aspects should be addressed, including insufficient assessment of test assumptions, lack of clarity in reporting software and effect sizes, vague descriptions of statistical methods, inappropriate use of post hoc tests, underappreciation of outliers or missing data, and misinterpretation of results (Ordak, 2025). The author proposes five initiatives to address these limitations and enhance transparency and reliability: Incorporate the SAMPL (Statistical Analyses and Methods in the Published Literature) guidelines into author instructions, require author declarations in cover letters, organise online training for editorial boards, advocate for the SAMPL guidelines at international events, and promote active engagement with the SAMPL guidelines (Ordak, 2025). Introduced in 2015, the SAMPL guidelines provide general principles for improving the reporting of statistical methods and results, several of which are summarised in Supplementary Figure 1. Readers are, however, encouraged to consult the article for a more comprehensive understanding (Lang and Altman, 2015).

EXCLI Journal 2026;25:1583-1587 – ISSN 1611-2156
Received: September 30, 2026, accepted: October 02, 2026, published: October 21, 2026

Reporting probability (*P*) values alone is insufficient for reanalysis, does not convey information about the magnitude of an effect, and is known to be influenced by sample size (Lang and Altman, 2015; Ordak, 2025; Tomczak and Tomczak, 2014). In particular, the standard error becomes smaller and the *P* value tends to decrease as the sample size increases, thereby increasing the probability of rejecting the null hypothesis ($H_0$) (Tomczak and Tomczak, 2014). One of the principles from the American Statistical Association's 2016 statement is that scientific conclusions should not be based solely on whether a *P* value passes a specific significance criterion (α) threshold (Wasserstein and Lazar, 2016). Alternatively, *P* values can be omitted from reports if point estimates and confidence intervals (CIs) are used (Griffiths and Needleman, 2019). This article highlights gaps in statistical reporting of experimental cancer therapeutics.

***Null hypothesis significance testing-related components***

Four key variables in statistical testing are sample size (*N*), α, effect size, and statistical power (Supplementary Table 1). In research planning, the required *N* to achieve a specified power for a given α and effect size can be calculated. For any statistical test, power can be determined with known *N*, α, and effect size. Notably, *N* increases as the required power increases, and *N* decreases as the effect size or α increases (Cohen, 1992). Overall, sufficient *N* is necessary for a Type I error (α) rate of 0.05 or 0.01 while achieving a statistical power of 0.80 or 0.90 (Supplementary Figure 2) (Serdar et al., 2021).

***Effect size estimates: From peer evaluation and the existing literature***

Effect size measures include the standardised mean difference (Cohen's *d*), Cohen's *f*, correlation coefficient (*r*), odds ratio, risk ratio, and eta squared ($\eta^2$) (Serdar et al., 2021). Cohen introduced small, medium, and large benchmarks for interpreting effect sizes across different statistical analyses (Cohen, 1992). For example, small, medium, and large effects are defined as Cohen's $d = 0.20$, 0.50, and 0.80 for independent means, and Cohen's $f = 0.10$, 0.25, and 0.40 for one-way analysis of variance (ANOVA). Whilst these effect size cut-offs may be useful for research with no prior knowledge, no universal rules apply. Rather, interpretations should consider the study's outcomes and comparisons with prior related studies (Vacha-Haase and Thompson, 2004), as well as the real-world implications of the observed effects (Griffiths and Needleman, 2019). Further reading can be found in several other sources, including (Ferguson, 2009; Fern and Monroe, 1996; Prentice and Miller, 1992).

Effect size estimates reflect the strength of relationships between variables, allowing comparison of results across studies (Tomczak and Tomczak, 2014). Unlike *P* values, which are influenced by both sample size and effect size, effect size estimates together with narrow 95 % CIs provide a more reliable metric of clinical or biological importance (Gan, 2026). In particular, as the effect size decreases, the power of a study decreases, and with very small sample sizes, only very large effects can be detected reliably; otherwise, studies are underpowered to reject the $H_0$ (Supplementary Figure 2) (Serdar et al., 2021). Large effect sizes that are not statistically significant may suggest the need for future research with greater power, whereas trivially small effect sizes that are significant because of a large sample size may prevent overinterpretation of the observed effects (Fritz et al., 2012).

Our previous peer review feedback recommended reporting effect size estimates, such as partial eta-squared ($\eta_p^2$) for ANOVA and Hedges's *g* for pairwise contrasts. These effect size measures can be obtained using software such as the Statistical Package for the Social Sciences (SPSS) or calculated manually as follows:

EXCLI Journal 2026;25:1583-1587 – ISSN 1611-2156
Received: September 30, 2026, accepted: October 02, 2026, published: October 21, 2026

$$\eta_p^2 = \frac{SS_{effect}}{SS_{effect} + SS_{error}} \text{ or } \eta_p^2 = \frac{F_{effect} \times df_{effect}}{(F_{effect} \times df_{effect}) + df_{error}}$$

$$\text{Hedges's } g = \text{Cohen's d} \times \left(1 - \frac{3}{4\,(N) - 9}\right)$$

Where SS (sum of squares), df (degrees of freedom), and $N = n_1 + n_2$ (Lakens, 2013).

Next, a literature review for 2025 was carried out using the Scopus database. Of the 197 search results, the first 150 articles in the exported list were reviewed sequentially. Of the 129 accessible full-text articles, 95 contained statistical analyses. Among the included studies, most focused on compound synthesis and drug formulation development, with *in vitro* and/or *in vivo* biological evaluation using various tumour models, including cell lines and mouse xenografts. Overall, although ANOVA and *t* tests are common, effect size estimates remain underreported in preclinical anticancer studies (Supplementary Table 2 and Supplementary Data). Data variability was most frequently reported as mean ± SD rather than mean (SD), although some used both SD and SE, and did not specify. Likewise, several studies reported both $P <$ and exact $P$ values, with the former predominating. Crucially, the lack of effect size estimates, 95 % CIs, and exact $P$ values reporting may contribute to their continued omission, as later studies may follow or refer to the existing literature. The assessment of test assumptions was generally not described. It was also observed that a number of studies reported statistically significant results ($P < 0.05$) but did not specify the statistical test used (marked by * in the Supplementary Data).

In a review of animal studies published in European Journal of Pharmacology, PLOS Biology, Pharmacological Research, Biomedicine and Pharmacotherapy, and British Journal of Pharmacology, 99 % (95 % CI: 95–100) relied exclusively on statistical hypothesis testing or $P$ values (Sharifan, 2026). In a different discipline, a review of articles published in the 2009 and 2010 volumes of the Journal of Experimental Psychology: General found that 58 % reported at least one measure of effect size. The type of effect size estimate generally depended on the statistical analysis performed; for mean differences, $\eta_p^2$ was most frequently reported for ANOVAs and Cohen's *d* for *t* tests (Fritz et al., 2012). Among 328 articles published in a selection of physiology journals in June 2017, 277 performed either ANOVA or *t* tests, whereas 127 used both but lacked essential information to verify the results (Weissgerber et al., 2018). Additionally, in a review of psychological research publications, effect size was negatively correlated with sample size ($r = -0.54$ [95 % CI: −60, −50]), but independent of the statistical test used (Kühberger et al., 2014).

In most studies, a $P$ value below the pre-specified $\alpha = 0.05$ is taken as evidence against $H_0$. By convention, a $P$ value greater than 0.05 is classified as not significant. However, statistical significance is not equivalent to clinical or biological importance, and a non-significant result indicates the absence of evidence for a difference, not that no difference exists (Altman and Bland, 1995). Moreover, statistical significance does not indicate that a result will be replicated, and the only way to generalise findings is through successful replication (Cohen, 1994). The American Statistical Association defines a $P$ value as the probability under a specified statistical model that a statistical summary of data would be equal to or more extreme than its observed values (e.g., the sample mean difference between two groups) (Wasserstein and Lazar, 2016). Notably, one journal has prohibited the reporting of $P$ values, statements of statistical significance, and CIs since 2014, and instead requires comprehensive descriptive statistics, including effect sizes (Trafimow and Marks, 2015). Further reading on this topic includes (Altman and Bland, 1995; Greenland et al., 2016; Griffiths and Needleman, 2019; Wasserstein and Lazar, 2016).

In summary, the findings and points emphasised may not be generalisable to other types of preclinical cancer research, such as genetic screens, drug interactions, and basic mechanistic studies. Where applicable, reporting effect size estimates with 95 % CIs provides information on the magnitude of interventional effects. From our perspective, the use of statistical tests and proper reporting can facilitate evidence synthesis and decisions regarding clinical translation. As improving research reporting quality is a shared effort, this piece hopefully helps promote it within the preclinical anticancer community. Indeed, to truly advance knowledge, the basics must be mastered, and knowledge without practice is like a seed never planted—full of potential yet bearing no fruit.

***Acknowledgments***

The authors thank Universiti Sains Malaysia for the Postdoctoral Fellowship.

***Conflict of interest***

The authors declare that they have no known conflicts of interest.

***Funding***

Not applicable.

***Consent for publication***

Not applicable.

***Data availability***

Not applicable.

***Ethics approval and consent to participate***

Not applicable.

***Artificial Intelligence (AI) – assisted technology***

AI-assisted illustration tools such as Gemini Notebook and Microsoft PowerPoint AI features were used to create figures.

***Author contributions***

SNH: Wrote, edited, and reviewed the paper. WMAWA: Edited and reviewed.

## REFERENCES

Altman DG, Bland JM. Absence of evidence is not evidence of absence. BMJ. 1995;311:485. https://doi.org/10.1136/bmj.311.7003.485

Cohen J. A power primer. Psychological bulletin. 1992;112:155–9. https://doi.org/10.1037/0033-2909.112.1.155

Cohen J. The earth is round (p < .05). American Psychologist. 1994;49:997–1003. https://doi.org/10.1037/0003-066X.49.12.997

Ferguson CJ. An effect size primer: A guide for clinicians and researchers. Professional Psychology: Research and Practice. 2009;40:532–8. https://doi.org/10.1037/a0015808

Fern EF, Monroe KB. Effect-Size Estimates: Issues and Problems in Interpretation. Journal of Consumer Research. 1996;23:89–105. https://doi.org/10.1086/209469

Fritz CO, Morris PE, Richler JJ. Effect size estimates: current use, calculations, and interpretation. Journal of experimental psychology General. 2012;141:2–18. https://doi.org/10.1037/a0024338

Gan W. Impact of Sample Size and Its Estimation in Medical Research. AJPM Focus. 2026;5:100451. https://doi.org/10.1016/j.focus.2025.100451

Greenland S, Senn SJ, Rothman KJ, Carlin JB, Poole C, Goodman SN, et al. Statistical tests, P values, confidence intervals, and power: a guide to misinterpretations. European Journal of Epidemiology. 2016;31:337–50. https://doi.org/10.1007/s10654-016-0149-3

Griffiths P, Needleman J. Statistical significance testing and p-values: Defending the indefensible? A discussion paper and position statement. International Journal of Nursing Studies. 2019;99:103384. https://doi.org/10.1016/j.ijnurstu.2019.07.001

Kühberger A, Fritz A, Scherndl T. Publication Bias in Psychology: A Diagnosis Based on the Correlation between Effect Size and Sample Size. PLOS ONE. 2014;9:e105825. https://doi.org/10.1371/journal.pone.0105825

Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAs. Frontiers in psychology. 2013;4:863. https://doi.org/10.3389/fpsyg.2013.00863

Lang TA, Altman DG. Basic statistical reporting for articles published in Biomedical Journals: The "Statistical Analyses and Methods in the Published Literature" or the SAMPL Guidelines. International Journal of Nursing Studies. 2015;52:5–9.
https://doi.org/10.1016/j.ijnurstu.2014.09.006

Ordak M. Implementation of SAMPL guidelines: Recommendations for improving statistical reporting in biomedical journals. Clinical Medicine. 2025;25: 100304.
https://doi.org/10.1016/j.clinme.2025.100304

Prentice DA, Miller DT. When small effects are impressive. Psychological bulletin. 1992;112:160–4.
https://doi.org/10.1037/0033-2909.112.1.160

Serdar CC, Cihan M, Yücel D, Serdar MA. Sample size, power and effect size revisited: simplified and practical approaches in pre-clinical, clinical and laboratory studies. Biochemia medica. 2021;31:010502.
https://doi.org/10.11613/BM.2021.010502

Sharifan A. From "No Difference" to "Trending": misinterpretation of statistical analysis in animal studies. Naunyn-Schmiedeberg's Archives of Pharmacology. 2026;399:1537–40.
https://doi.org/10.1007/s00210-025-04440-0

Tomczak M, Tomczak E. The need to report effect size estimates revisited. An overview of some recommended measures of effect size. TRENDS in Sport Sciences. 2014;21:19–25.

Trafimow D, Marks M. Editorial. Basic and Applied Social Psychology. 2015;37:1–2.
https://doi.org/10.1080/01973533.2015.1012991

Vacha-Haase T, Thompson B. How to estimate and interpret various effect sizes. Journal of Counseling Psychology. 2004;51:473–81.
https://doi.org/10.1037/0022-0167.51.4.473

Wasserstein RL, Lazar NA. The ASA Statement on p-Values: Context, Process, and Purpose. The American Statistician. 2016;70:129–33.
https://doi.org/10.1080/00031305.2016.1154108

Weissgerber TL, Garcia-Valencia O, Garovic VD, Milic NM, Winham SJ. Why we need to report more than 'Data were Analyzed by t-tests or ANOVA'. eLife. 2018;7:e36163.
https://doi.org/10.7554/eLife.36163