Reviewing research reporting in randomised controlled trials: Confidence and *P*-values

INTRODUCTION

Despite the significant strides made by journals and funding agencies in enhancing research reporting standards, a notable gap persists in the field of anaesthesia research about statistical tools. To bridge this gap, we undertook a cross-sectional study of all the randomised controlled trials published in 2023 in the Indian Journal of Anaesthesia (IJA). Our study focused on reporting confidence intervals (CIs) and *P* values, a topic that has not been extensively explored in this context.[1]

METHODS

We conducted a cross-sectional analysis of research papers published in the Indian Journal of Anaesthesia from January 2023 to December 2023. We (NS and VG) independently searched each monthly issue on the IJA official website, screened the titles and abstracts, and populated the randomised controlled trials that met the inclusion criteria.

Our exploration of statistical reporting elements was based on a comprehensive questionnaire, partially derived from the Consolidated Standards of Reporting Trials (CONSORT) 2010 standards. This questionnaire [Table 1] was rigorously pilot-tested on 20% of the RCTs, with each trial examined independently over 2 weeks. Any disagreements were resolved through discussion, ensuring the reliability of our findings. Our assessment focused on key statistical reporting areas, including baseline data, imbalance assessment, and outcome reporting as CIs. These areas are crucial for the interpretation and replication of research findings.

As we aimed to include and describe all the randomised studies in the past year, no formal sample size calculation was done. All analyses were performed using Google Sheets and R Core Team (2023): R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing, Vienna, Austria (https://www.R-project.org/), and Rversion 4.3.2 ('eye holes').

RESULTS

The last search date was 21st January 2024, and candidate studies were retrieved from the IJA website. All 50 included studies were parallel-group trials [Figure 1].

Baseline demographic data were compared using *P*-values in 58% of the trials. However, the frequency of trials reporting these decreased over the months [Figure 2b]. Regarding outcome reporting, CIs were reported for the primary and secondary outcomes in 24% of the studies. The proportion of studies reporting these increased over months [Figure 2c]. The trends of reporting characteristics are shown in Figure 2a-d.

DISCUSSION

Applying null hypothesis significance testing in comparing baseline variables using *P* values does not seem valid. For instance, if the mean age were 46 in the treatment group and 52 in the control group with $P < 0.05$, we would reject the null hypothesis of no difference in age, and one may read the alternate hypothesis as 'the participants were younger because they were allocated into the treatment group'.

Similarly, one should not give too much importance to only the outcomes with $P < 0.05$. To consider the results/effects significant only if $P < 0.05$ is too simplistic and the term 'significant' loses its semantic meaning when clinically significant treatment effects
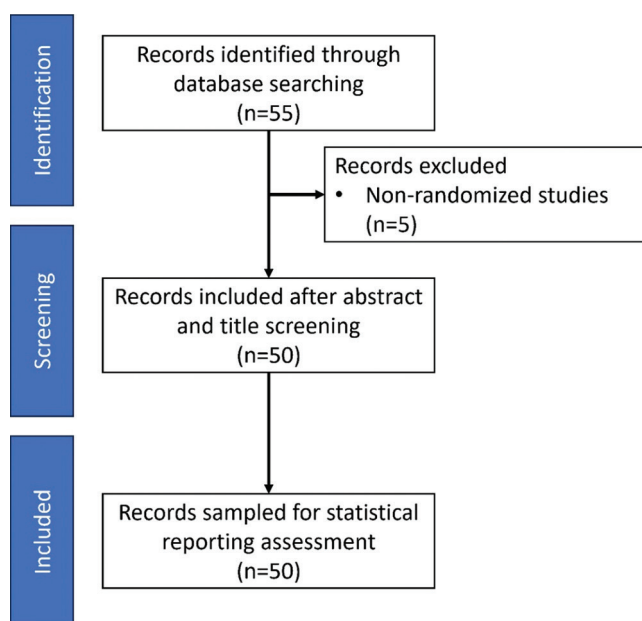

Figure 1: Study flow. n=number of patients

Table 1: Summary of the results of the questionnaire assessing reporting characteristics	
Question	**Summary statistic (*n*=50) [95% CI]**
Was the primary outcome mentioned in the title? (Yes/No)	25 (50%) [37%, 63%]
Was the trial design mentioned in the title? (Yes/No)	49 (98%) [88%, 100%]
Was a primary outcome specified? (Yes/No)	
Yes	36 (72%) [57%, 83%]
Poorly defined	13 (26%) [15%, 41%]
No	1 (2.0%) [0.10%, 12%]
Sample size calculation	
Was the sample size calculation based on the primary outcome? (Yes/No)	45 (90%) [77%, 96%]
Was a target effect size mentioned? (Yes/No)	35 (70%) [55%, 82%]
What is the source of the control group's target effect size/effect? Is it arbitrary/single study/multiple studies?	
Single study	27 (54%) [39%, 68%]
Pilot	14 (28%) [17%, 43%]
Multiple studies	4 (8.0%) [2.6%, 20%]
Arbitrary	1 (2.0%) [0.10%, 12%]
Meta-analysis	1 (2.0%) [0.10%, 12%]
Retrospective audit	1 (2.0%) [0.10%, 12%]
Sample size not calculated	1 (2.0%) [0.10%, 12%]
Rule of thumb	1 (2.0%) [0.10%, 12%]
Was a software/formula used to calculate the sample size mentioned? (Yes/No)	17 (34%) [22%, 49%]
Was a specific statistical test mentioned against which the sample size is being calculated? (Name of the test/No)	
No	45 (90%) [77%, 96%]
Formula	1 (2.0%) [0.10%, 12%]
Non-inferiority for the mean difference	1 (2.0%) [0.10%, 12%]
One-tailed *t*-test	1 (2.0%) [0.10%, 12%]
Two sample proportion test	1 (2.0%) [0.10%, 12%]
Unpaired *t*-test	1 (2.0%) [0.10%, 12%]
Were enough details provided to replicate the sample size calculation? (Yes/No)	7 (14%) [6.3%, 27%]
What was the sample size? Median (IQR) [CI for median]	75 (60, 92) [60, 90]
Statistical analysis	
Were the statistical methods used to compare primary outcomes specifically mentioned? (Yes/No)	17 (34%) [22%, 49%]
Was the primary outcome analysed as per the method targeted in the sample size calculation?	
Detail insufficient	37 (74%) [59%, 85%]
No	7 (14%) [6.3%, 27%]
Yes	6 (12%) [5.0%, 25%]
Were correction factors (e.g. Bonferroni) used in multiple testing instances?	
No	37 (74%) [59%, 85%]
Yes	9 (18%) [9.0%, 32%]
Not Applicable	4 (8.0%) [2.6%, 20%]
Was a participant flow diagram presented? (Yes/No)	48 (96%) [85%, 99%]
Were appropriate measures of central tendency used to represent the data? (e.g. median IQR for ordinal outcomes, mean SD for continuous outcomes, etc.) (Yes/No)	37 (74%) [59%, 85%]
Were baseline data compared using P values? (Yes/No)	29 (58%) [43%, 72%]
Outcome reporting	
Were confidence intervals reported for primary outcomes? (Yes/No)	12 (24%) [14%, 38%]
Were confidence intervals reported for secondary outcomes? (Yes/No)	12 (24%) [14%, 38%]
Were effect size estimates (Differences in mean/median/proportion) and their CIs reported? (Yes/No)	12 (24%) [14%, 38%]
Were the outcomes represented by appropriate graphs? (Yes/No)	7 (14%) [6.3%, 27%]
Were sub-group analyses, if any, pre-specified? - no sub-group analysis/pre-specified/post hoc	
No sub-group analysis	47 (94%) [82%, 98%]
Pre-specified	3 (6.0%) [1.6%, 18%]
Was the final interpretation based on the intended primary outcome?	34 (68%) [53%, 80%]

Data expressed as numbers (percentages) (95% confidence interval). For 'Yes/No' questions the proportions representing 'Yes' have been provided.
CI – Confidence intervals, IQR-interquartile range, SD-standard deviation, n=number of patients

are ignored as soon as $P = 0.05$. For example, trialists exploring early airway pressure release ventilation against low tidal volume ventilation in 138 patients with ARDS had concluded that the ICU mortality was similar between the groups even when there was an absolute risk difference of -14.6% ($P = 0.053$).[2] One
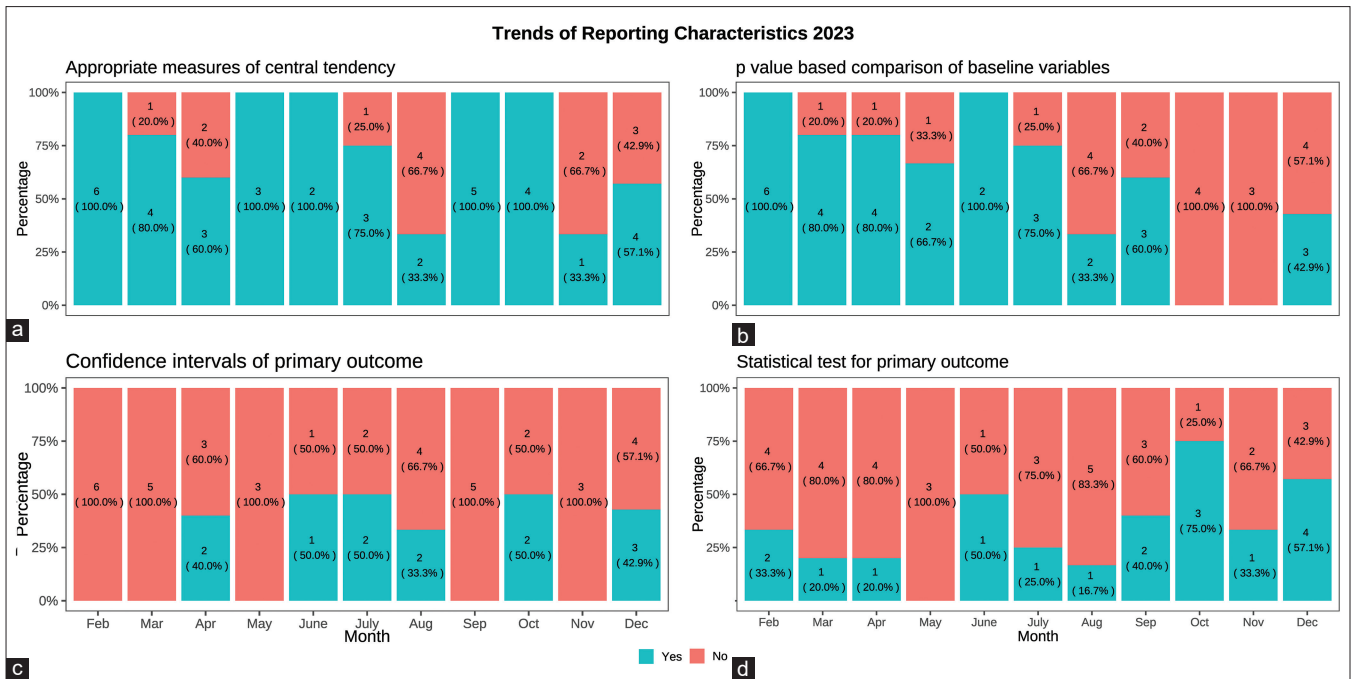
Figure 2: Trends of reporting characteristics. (a) Appropriate measures of central tendency, (b) *P*-value based comparison of baseline variables, (c) Confidence interval of primary outcome, (d) Statistical test for primary outcome

cannot find a logical reason to explain why a *P* value of 0.053 would be less significant than a *P*-value of 0.049, especially in the face of hard outcomes such as mortality with a clinically important difference. A *P* > 0.05 does not mean that the null hypothesis is true; it only means that we could not find sufficient evidence to reject the null hypothesis.[3,4] The absence of evidence is not the same as evidence of absence.

Reporting confidence intervals (CIs) for the difference between treatment arms can help distinguish statistical significance from clinical or practical importance.[3,5,6] The CI denotes the range of values within which an estimate/population parameter is anticipated to fall a specified percentage (the confidence level or coverage level) of the time when the experiment is replicated, or the population is re-sampled using the same methodology. In a 95% CI, there is a 95% probability that the range of the interval derived from the sample contains the true value of the population parameter.[3] The authors of RCTs in IJA in 2023 have reported CIs for primary and/or secondary outcomes and CIs for the effect size estimates.[7-18] The awareness of the relevance of CIs and the willingness to adopt CI reporting are indeed encouraging. However, it is important to understand that merely reporting the CIs is not enough; interpreting results based on CIs, especially the CI of between-group differences (treatment effects), is essential. This should include explicitly interpreting the width of the CI and not restricting oneself to merely commenting if it consists of the null. This will improve the quality of research reporting and facilitate decision-making.

The American Statistical Association released a 'Statement on Statistical Significance and *P*-values' with six principles.[19] A few editorials have explained the same, which has also made its way into current major textbooks.[20,21] Current reporting checklists, such as the CONSORT checklist, also require confidence intervals to be reported.[22]

CONCLUSION

There is a massive commitment in terms of time and resources when we perform clinical research and peer review each other's work. As physician-scientists, we need to strengthen our understanding of statistical concepts involved in conducting our research and ensure robust reporting of the same. There appears to be an improving trend in reporting the statistical elements over the months in the RCTs published in the IJA, especially in providing confidence intervals for effect estimates. In addition, placing excessive reliance on *P* values alone should be discouraged.

Financial support and sponsorship
Nil.

Conflicts of interest

There are no conflicts of interest.

ORCID

Venkata Ganesh: https://orcid.org/0000-0001-8265-5198

Neeru Sahni: https://orcid.org/0000-0002-8506-8076

Venkata Ganesh, Neeru Sahni

Department of Anaesthesia and Intensive Care, PGIMER,
Chandigarh, India

Address for correspondence:
Dr. Neeru Sahni,
Department of Anaesthesia and Intensive Care, Postgraduate
Institute of Medical Education and Research, Chandigarh - 160 012,
India.
E-mail: neerunalin@yahoo.com

REFERENCES

1. Diong J, Butler AA, Gandevia SC, Héroux ME. Poor statistical reporting, inadequate data presentation and spin persist despite editorial advice. PLoS One 2018;13:e0202121. doi: 10.1371/journal.pone.0202121
2. Zhou Y, Jin X, Lv Y, Wang P, Yang Y, Liang G, *et al.* Early application of airway pressure release ventilation may reduce the duration of mechanical ventilation in acute respiratory distress syndrome. Intensive Care Med 2017;43:1648-59.
3. Analysing Means and Proportions. In: Armitage P, Berry G, Matthews JNS, editors, Statistical Methods in Medical Research. 4th ed. (reprint) New Delhi, India: Wiley Blackwell; 2017. p. 88-90.
4. Gopinath R. To reveal or to conceal: Appropriate statistical analysis is a moral obligation for authors in modern medicine. Indian J Anaesth 2023;67:323-5.
5. Altman DG, Bland JM. How to obtain the confidence interval from a *P* value. BMJ 2011;343:d2090. doi: 10.1136/bmj.d2090
6. du Prel J-B, Hommel G, Röhrig B, Blettner M. Confidence interval or *P* value?: Part 4 of a series on evaluation of scientific publications. Dtsch Arztebl Int 2009;106:335-9.
7. Sharma L, Bhatia P, Mohammed S, Sethi P, Chhabra S, Kumar M. Comparison of erector spinae plane block and thoracic paravertebral block for postoperative analgesia in patients undergoing modified radical mastectomy: A randomised controlled non-inferiority trial. Indian J Anaesth 2023;67:357-63.
8. Chhabra A, Dave M, Jeenger L, Meena R, Aggarwal I, Partani S. Comparison of Quality of Recovery (QoR-15) following the administration of intravenous lignocaine and fentanyl in patients undergoing septoplasty under general anaesthesia: A double-blinded, randomised, controlled trial. Indian J Anaesth 2023;67:388-93.
9. Balakrishnan A, Chhabra A, Kumar A, Talawar P, Bhoi D, Garg H. Comparison of continuous transmuscular quadratus lumborum block and continuous psoas compartment block for posterior total hip arthroplasty: A randomised controlled trial. Indian J Anaesth 2023;67:530-6.
10. Singh S, Avinash R, Jaiswal S, Kumari A. Comparison of safety and efficacy of thoracic epidural block and erector spinae plane block for analgesia in patients with multiple rib fractures: A pilot single-blinded, randomised controlled trial. Indian J Anaesth 2023;67:614-9.
11. Gupta K, Gupta M, Sabharwal N, Subramanium B, Belani KG, Chan V. Ultrasound-guided anterior suprascapular nerve block versus interscalene brachial plexus block for arthroscopic shoulder surgery: A randomised controlled study. Indian J Anaesth 2023;67:595-602.
12. Ahuja S, Kaur G, Garg K, Grewal A. Conventional versus reverse insertion of i-gel® in overweight and obese patients – Interventional randomised controlled trial. Indian J Anaesth 2023;67:708-13.
13. Ranjan V, Singh S. Comparison of ultrasound-guided transversus abdominis plane block and caudal epidural block for postoperative analgesia in paediatric lower abdominal surgeries: A randomised controlled trial. Indian J Anaesth 2023;67:720-4.
14. Shetmahajan M, Kamalakar M, Narkhede A, Bakshi S. Analgesic efficacy of the inferior alveolar nerve block for maxillofacial cancer surgery under general anaesthesia – A randomised controlled study. Indian J Anaesth 2023;67:880-4.
15. Eid GM, El said Shaban S, Mostafa TA. Comparison of ultrasound-guided genicular nerve block and knee periarticular infiltration for postoperative pain and functional outcomes in knee arthroplasty – A randomised trial. Indian J Anaesth 2023;67:885-92.
16. Khot PP, Desai SN, Bale SP, Aradhya BN. Comparison of ultrasound-guided paravertebral block versus erector spinae plane block for postoperative analgesia after percutaneous nephrolithotomy – A randomised, double-blind, controlled study. Indian J Anaesth 2023;67:1110-5.
17. Mohta M, Mounika TB, Tyagi A. Effect of timing of intraoperative administration of paracetamol on postoperative shivering: A randomised double-blind controlled trial. Indian J Anaesth 2023;67:1071-6.
18. Avci O, Gundogdu O, Balci F, Tekcan MN, Ozbey M. Efficacy of serratus posterior superior intercostal plane block (SPSIPB) on post-operative pain and total analgesic consumption in patients undergoing video-assisted thoracoscopic surgery (VATS): A double-blinded randomised controlled trial. Indian J Anaesth 2023;67:1116-22.
19. Wasserstein RL, Lazar NA. The ASA statement on *P* values: Context, process, and purpose. Am Stat 2016;70:129-33.
20. Yaddanapudi LN. The American Statistical Association's statement on *P* values is explained. J Anaesthesiol Clin Pharmacol 2016;32:421.
21. Whitlock EL, Chen CL. Interpreting the Medical Literature. In: Gropper AM, editor. Mill Anesth. 9th ed. Philadelphia, PA, USA: Elsevier; 2020. p. 2820.
22. Moher D, Hopewell S, Schulz KF, Montori V, Gøtzsche PC, Devereaux PJ, *et al.* CONSORT 2010 Explanation and Elaboration: Updated guidelines for reporting parallel group randomised trials. BMJ 2010;340:c869. doi: 10.1136/bmj.c869

Access this article online

Quick response code

Website: https://journals.lww.com/ijaweb

DOI: 10.4103/ija.ija_189_24

How to cite this article: Ganesh V, Sahni N. Reviewing research reporting in randomised controlled trials: Confidence and *P* values. Indian J Anaesth 2024;68:492-5.