

Statistical Issues in the Oliceridine versus Sufentanil Trial for Elderly Thoracoscopic Surgery [Letter]

Dan Xu*, Hui Ma*, Dongnan Hou

Department of Anesthesiology, The Second Affiliated Hospital of Dalian Medical University, Dalian, People's Republic of China

*These authors contributed equally to this work

Correspondence: Dongnan Hou, Department of Anesthesiology, The Second Affiliated Hospital of Dalian Medical University, Dalian, People's Republic of China, Tel +86 17709872323, Email dongnanhouxy@163.com

Dear editor

Cai et al conducted a randomized double-blind trial comparing oliceridine with sufentanil for patient-controlled intravenous analgesia in 247 elderly patients undergoing thoracoscopic lung surgery.¹ The authors concluded that oliceridine increased the proportion of patients achieving satisfactory analgesia with minimal emesis (SAME) within the first three postoperative days (46.3% vs 32.3%; $P=0.023$) and provided better early recovery. The study addresses a clinically important topic, and the overall design is commendable. However, we would like to respectfully raise three methodological points that may help readers interpret the findings and inform future research.

Primary Analysis on the Per-Protocol Set Rather Than Intention-to-Treat

The authors state that “the per-protocol set (PPS) analysis was used to compare the SAME endpoint”. In randomised controlled trials, the intention-to-treat (ITT) principle is widely recommended because it preserves the benefits of randomisation and avoids bias due to post-randomisation exclusions.^{2,3} In this study, one patient in the oliceridine group was lost to follow-up and excluded from the PPS. Moreover, any protocol deviations (eg, early discontinuation of the patient-controlled analgesia pump due to nausea or inadequate pain relief) would also have been excluded from the PPS. If such deviations are related to treatment tolerability or efficacy, the PPS may overestimate the true treatment effect.

We suggest that the authors report the ITT estimate for the SAME endpoint as a sensitivity analysis. This could be done by including all randomised patients in their original groups and using a conservative imputation for the single lost patient (eg, considering that patient as not having achieved SAME). If the ITT result remains statistically significant and the point estimate is similar, confidence in the conclusion would be strengthened. If not, a more cautious interpretation would be warranted. This is not to criticise the authors' choice, but to align with current reporting standards that recommend ITT as the primary analysis and PPS as a supportive one.²

Sample Size Calculation Based on a Small Pilot Study

The sample size was calculated from a pilot study of only 30 patients (15 per group), which gave SAME rates of 60% for oliceridine and 40% for sufentanil – an absolute difference of 20%. Using these figures, the authors estimated that 99 patients per group would provide 80% power. In the final trial, the observed absolute difference was 14.1% (from 32.3% to 46.3%), which is smaller than anticipated. A post-hoc calculation based on the observed difference would likely yield power below 80%.

Small pilot studies often provide imprecise estimates of effect size; the 95% confidence interval around a 20% difference from only 15 patients per group would be extremely wide.⁴ Consequently, the actual trial may have been somewhat underpowered to detect the true effect. This is reflected in the width of the 95% confidence interval for the primary outcome (2.0% to 26.1%), which just excludes the null. We do not suggest that the study should be dismissed; rather, we recommend that the authors report a sensitivity analysis using different assumptions for the SAME definition (eg, a more stringent postoperative nausea and vomiting (PONV) threshold) or for handling of missing data, to evaluate how robust the conclusion is. Additionally, a post-hoc power calculation based on the observed effect could be presented for transparency. These additional analyses would help readers gauge the stability of the finding.

Multiplicity Issues in Secondary Outcomes

The study examined a considerable number of secondary outcomes, including numerical rating scale (NRS) pain scores on three separate days (POD1, POD2, POD3), 5-item Quality of Recovery (QoR-15) scores on three days, need for rescue analgesia, number of PCIA boluses, time to ambulation, and several safety events. For each of these, the authors performed separate tests at each time point and reported unadjusted *P* values (eg, NRS on POD1 *P*=0.018, NRS on POD2 *P*=0.039, QoR-15 on POD1 *P*<0.001, QoR-15 on POD2 *P*=0.010). No correction for multiple comparisons was applied.

Even for the primary SAME endpoint, comparisons were made on POD1, POD2, and POD3 (Table 2 in the original article). If a conservative Bonferroni correction were applied for these three tests ($\alpha=0.05/3\approx 0.0167$), the *P* value for SAME on POD2 (0.024) would no longer be statistically significant. For the secondary outcomes, the risk of at least one false-positive finding is even higher.⁵ We fully recognise that excessive correction may increase the risk of type II errors, especially in exploratory analyses. However, we believe that the authors could improve the paper by clearly stating which secondary outcomes were pre-specified (rather than post-hoc) and by applying a more balanced multiplicity adjustment, such as the Hochberg procedure or false discovery rate control. Alternatively, using a global test (eg, a repeated-measures model with a single group-by-time interaction) would reduce the number of separate comparisons. Reporting adjusted *P* values or providing the raw data for readers to perform their own adjustment would also be helpful.

Additional Minor Clarifications

A few other details would benefit from clarification. First, the definition of “within the first 3 postoperative days” – does it include postoperative day 0 (the day of surgery)? The authors mention that NRS was assessed at 8-hour intervals; how was the “daily average coughing NRS” calculated (arithmetic mean of three measurements, or median)? Second, the use of repeated-measures ANOVA for NRS scores, which are often skewed, requires checking the normality of residuals and sphericity; if sphericity was violated, appropriate corrections (Greenhouse-Geisser or Huynh-Feldt) should be reported. Third, the trial registration (ChiCTR2500102213) is provided, but the registration date is not given; it would be important to confirm that enrolment started after registration, as required by CONSORT.²

Conclusion

Cai et al have conducted a valuable study on an important topic. Our comments are intended as a constructive methodological discussion, not as a criticism of the work. We believe that providing ITT estimates, a sensitivity analysis for sample size assumptions, and a clearer handling of multiplicity would make the evidence even more robust. We thank the authors for their contribution and look forward to their response.

Disclosure

The authors report no conflicts of interest in this communication.

References

1. Cai Y, Jiang Y, Zhang Q, Yang J, Wang Z, Sun H. The satisfactory analgesia and minimal emesis of elderly patients after thoracoscopic lung surgery: oliceridine versus sufentanil in a randomized controlled trial. *Drug Des Devel Ther.* 2026;20:593306. doi:10.2147/DDDT.S593306
2. Schulz KF, Altman DG, Moher D; CONSORT Group. CONSORT 2010 statement: updated guidelines for reporting parallel group randomised trials. *BMJ.* 2010;340:c332. doi:10.1136/bmj.c332

3. ICH Harmonised Guideline. Addendum on estimands and sensitivity analysis in clinical trials to the guideline on statistical principles for clinical trials E9(R1). ICH; 2019.
4. Button KS, Ioannidis JPA, Mokrysz C, et al. Power failure: why small sample size undermines the reliability of neuroscience. *Nat Rev Neurosci*. 2013;14(5):365–376. doi:10.1038/nrn3475
5. Li G, Taljaard M, Van den Heuvel ER, et al. An introduction to multiplicity issues in clinical trials: the what, why, when and how. *Int J Epidemiol*. 2017;46(2):746–755. doi:10.1093/ije/dyw320

Dove Medical Press encourages responsible, free and frank academic debate. The content of the Drug Design, Development and Therapy 'letters to the editor' section does not necessarily represent the views of Dove Medical Press, its officers, agents, employees, related entities or the Drug Design, Development and Therapy editors. While all reasonable steps have been taken to confirm the content of each letter, Dove Medical Press accepts no liability in respect of the content of any letter, nor is it responsible for the content and accuracy of any letter to the editor.

Drug Design, Development and Therapy

Dovepress
Taylor & Francis Group

Publish your work in this journal

Drug Design, Development and Therapy is an international, peer-reviewed open-access journal that spans the spectrum of drug design and development through to clinical applications. Clinical outcomes, patient safety, and programs for the development and effective, safe, and sustained use of medicines are a feature of the journal, which has also been accepted for indexing on PubMed Central. The manuscript management system is completely online and includes a very quick and fair peer-review system, which is all easy to use. Visit <http://www.dovepress.com/testimonials.php> to read real quotes from published authors.

Submit your manuscript here: <https://www.dovepress.com/drug-design-development-and-therapy-journal>