Journal of Anesthesia (2024) 38:421–424
https://doi.org/10.1007/s00540-024-03355-z

EDITORIAL

Understanding three standard study designs for randomized controlled trials

Naoko Niimi[1] · Alan Yang[2] · Makoto Sumie[1,2,3,4] · Jason Hayes[1] · Jason T. Maynes[1,5] · Kazuyoshi Aoyama[1,2]

Received: 15 May 2024 / Accepted: 24 May 2024 / Published online: 24 June 2024

Keywords Study design · Randomized controlled trials · Pediatric anesthesia

Statistically significant difference is a widely used approach to interpret the result of randomized controlled trials (RCTs). However, there are several different study designs for RCTs and hypothesis, method of statistical testing, and calculation of sample size can differ depending on each study design. Misuse and misinterpretation of the study designs are frequently observed [1], which can lead to fragile RCTs.

There are mainly three study designs for RCTs. The first study design is a superiority trial, which aims to demonstrate a statistically significant difference between two interventions. The second design is a non-inferiority trial, which aims to demonstrate one intervention is 'no worse' than the other. The third study design is an equivalence trial which aims to demonstrate one intervention is 'no better and no worse' compared to the other. Figure 1 illustrates the interpretation of each trial design while their characteristics including the study hypothesis and examples in pediatric anesthesia trials are summarized in Table 1. Non-inferiority and equivalence trials are typically compared with existing standard interventions since conducting a superiority trial, which typically involves comparing against a placebo, may raise ethical concerns. In a superiority trial, rejecting the null hypothesis hinges on demonstrating a statistically significant difference between interventions, typically evaluated through *P*-values or confidence intervals. Therefore, a preset margin is unnecessary in a superiority trial. On the other hand, non-inferiority and equivalence trials prioritize establishing a pre-analysis margin (denoted as $\pm\Delta$) that defines the largest acceptable difference between interventions. This margin, critical for determining whether to reject the null hypothesis, is often informed by prior studies. If the differences between the mean response of an outcome is larger than this predetermined margin, the proposed intervention is deemed to have failed in rejecting the null hypothesis. Mean difference (MD) is commonly used to compare the mean responses of an outcome between groups; however, other measures can also be utilized to assess differences between groups such as odds ratios and risk ratios [2]. Additionally, superior and equivalence trials require two-sided confidence intervals (CIs) to assess both potential equivalence and improvement whereas non-inferiority trials use one-sided CI to demonstrate non-inferiority, focusing on a single direction to ensure the proposed intervention is not significantly worse.

Recently, Cho et al. studied their hypothesis that a 2 h fasting period is not inferior to a 4 h fasting period when evaluating the gastric cross-sectional area in the right lateral decubitus position (CSA_{RLDP}) of pediatric patients undergoing echocardiography [6]. In calculating sample size for a non-inferiority trial, power, significance level, and non-inferiority margin are important components to be determined during the trial's design and planning phases [7]. Power and significance level relate to two different types of errors in hypothesis testing. Significance level is defined as the probability of rejecting a null hypothesis when it is true, in other words, the probability of making a Type I error. Power is considered as a probability of correctly rejecting

✉ Kazuyoshi Aoyama
kazu.aoyama@utoronto.ca

1 Department of Anesthesia and Pain Medicine, The Hospital for Sick Children, 555 University Ave, #2211, Toronto, ON M5G 1X8, Canada

2 Program in Child Health Evaluative Sciences, SickKids Research Institute, Toronto, ON, Canada

3 Department of Anesthesiology, St. Mary's Hospital, Fukuoka, Japan

4 Department of Anesthesiology and Critical Care Medicine, Graduate School of Medical Sciences, Kyushu University, Fukuoka, Japan

5 Program in Molecular Medicine, SickKids Research Institute, Toronto, ON, Canada

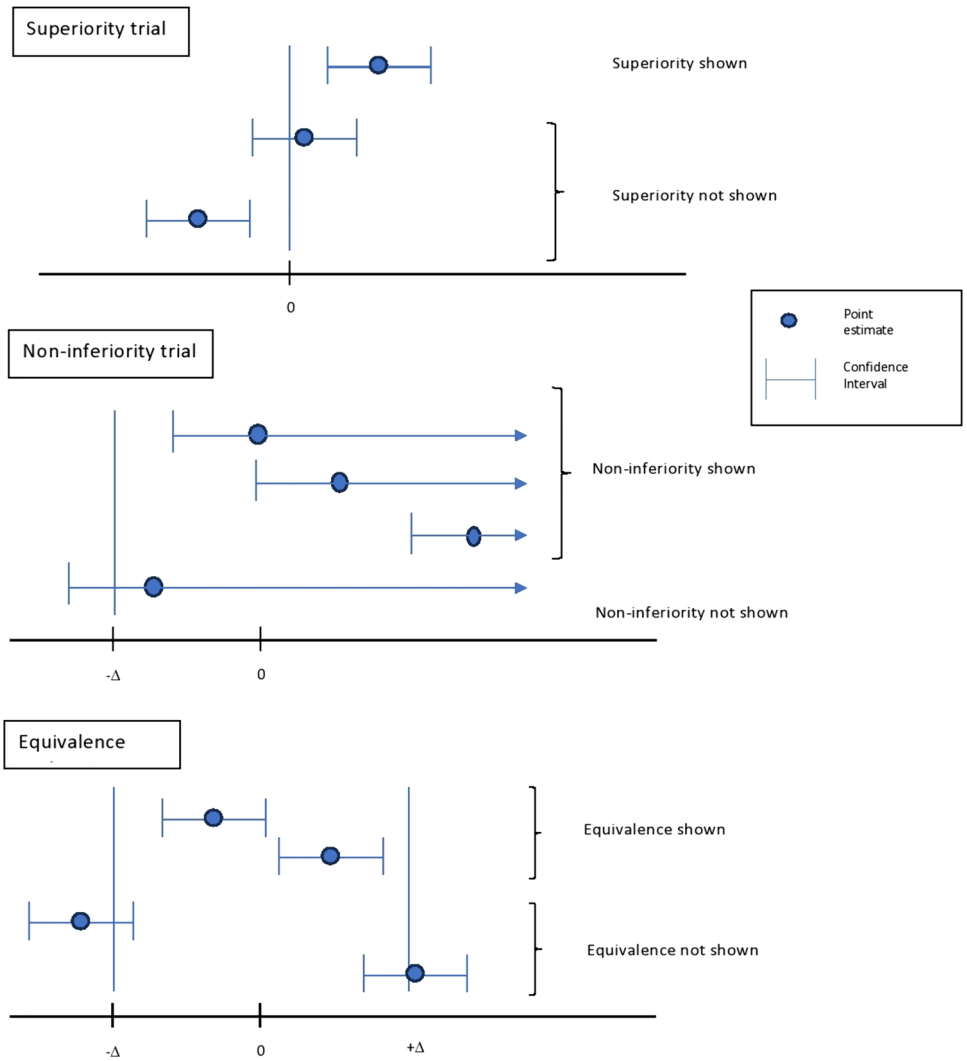

Fig. 1 Interpretations of results in superiority, non-inferiority, and equivalence trials with various confidence intervals and preset margins of the trial. The black horizontal line represents the mean difference (MD) between two groups (e.g., new treatment versus existing treatment), the dot represents the point estimate computed from the trial, and the error bar represents the confidence interval (CI) of the point estimate. These are fundamental to understanding the trial results. This figure evaluates the result in MD, so one can suppose a positive value (>0) can be interpreted as the new treatment being better. When one can suppose the trial explores risk as harm in a relative form, such as risk ratio (RR); the new treatment is considered better when the CI is between 0 and 1. Superiority can be declared if the CI is above MD of 0 (or less RR of 1); therefore, a preset margin is unnecessary in a superiority trial. The non-inferiority margin is demonstrated as $-\Delta$, and non-inferiority is shown when the CI of MD is above $-\Delta$. In a non-inferiority trial, the CI is one-sided, with the right side extended indefinitely to the edge of the graph. Similarly, the equivalence margin is demonstrated as the region between $-\Delta$ and $+\Delta$, where equivalence is shown when the CI of MD is between this margin.

the null hypothesis when it is false. In other words, power can be considered as the probability of avoiding a Type II error, which is not rejecting the null hypothesis when it is false. The authors determined the non-inferiority margin ($-\Delta$) as 2.1 cm^2, which was calculated by MD between grade 1 stomach (gastric contents exist but not to the extent to increase the risk of pulmonary aspiration) and grade 2 stomach (known to increase the risk of pulmonary aspiration) from a previous study [8]. However, since the population of the previous study slightly differs from the current study's population, this difference undermines the justification of the chosen non-inferiority margin. Ideally, population characteristics such as age groups and medical conditions between the current study and the study the authors used to determine the non-inferiority margin should be sufficiently similar.

Anesthesiology trials are known to be fragile and often underpowered both in adult and pediatric populations [9–11]. Underpowered trials are often impacted by a small number of participants and adverse event rates. One

Table 1 Examples of pediatric anesthesia trials and their characteristics in each trial design

	Examples in pediatric anesthesia trials	PICO question	Study hypothesis
Superiority trial	Prospective, randomized, double-blind, double-dummy, active-controlled, phase 3 clinical trial comparing the safety and efficacy of intranasal dexmedetomidine to oral midazolam as premedication for propofol sedation in pediatric patients undergoing magnetic resonance imaging: the MIDEX MRI trial Wabelo et al. [3]	P: Children aged 6 months to 6 years undergoing elective MRI I: Intranasal dexmedetoidine C: Oral midazolam O: Adverse event (i.e. bradycardia, hypotension, desaturation, involuntary movements)	Intranasal dexmedetomidine is superior to oral midazolam as premedication in reducing the occurrence of adverse events
Non-inferiority trial	Direct versus video laryngoscopy with standard blades for neonatal and infant tracheal intubation with supplemental oxygen: a multicentre, non-inferiority, randomised controlled trial Riva et al. [4]	P: Patient aged up to 52 weeks requiring tracheal intubation I: Intubation with direct laryngoscopy C: Intubation with video laryngoscopy O: Proportion of first-attempt tracheal intubation success	Direct laryngoscopy is not inferior to video laryngoscopy for tracheal intubation
Equivalence trial	Articles Neurodevelopmental outcome at 5 years of age after general anaesthesia or awake-regional anaesthesia in infancy (GAS): an international, multicentre, randomised, controlled equivalence trial Mccann et al. [5]	P: Infants less than 60 weeks undergoing inguinal herniorrhaphy I: Awake regional anesthetic C: Sevoflurane-based general anesthetic O: FSIQ on the WPPSI-III at 5 years of age	There are no differences in neurodevelopmental outcome between general anesthesia and regional anesthesia

MRI magnetic resonance imaging, *P* population, *I* intervention, *C* comparison, *O* outcomes, *FSIQ* full-scale intelligence quotient, *WPPSI-III* Wechsler Preschool and Primary Scale of Intelligence third edition

contributing factor to the small participant count could be an inaccurate calculation of the required sample size. Therefore, conducting proper power and sample size calculations corresponding to each study design is a vital part to improving the quality of these trials [12–15].

Acknowledgements JTM and KA acknowledge the Department of Anesthesiology and Pain Medicine, University of Toronto, and the Hospital for Sick Children, for secured academic time to conduct the current work as a recipient of a Merit Award 2023–2025.

Author contributions NN and KA wrote the initial and subsequent draft of the manuscript, which AY, MS, JH and JTM commented and edited. All authors approved the final version.

Funding This work was supported by the Canadian Anesthesiologists' Society Research Award 2022–2024 (KA) and Project Grants (PJX179857, PJT183603) 2022–2025, Canadian Institutes of Health Research (KA).

Declarations

Conflict of interest The authors declare that they have no competing interests related to this publication.

References

1. Greene CJ, Morland LA, Durkalski VL, Frueh BC. Noninferiority and equivalence designs: issues and implications for mental health research. J Trauma Stress. 2008. https://doi.org/10.1002/jts.20367.
2. Wang B, Wang H, Tu XM, Feng C. Biostatistics in psychiatry. Shanghai Arch Psychiatry. 2017;29:385–8. https://doi.org/10.11919/j.issn.1002-0829.217163.
3. Wabelo ON, Schmartz D, Giancursio M, De Pooter F, Caruso G, Fils JF, Van der Linden P. Prospective, randomized, double-blind, double-dummy, active-controlled, phase 3 clinical trial comparing the safety and efficacy of intranasal dexmedetomidine to oral midazolam as premedication for propofol sedation in pediatric patients undergoing magnetic resonance imaging: the MIDEX MRI trial. Trials. 2023;24(1):518. https://doi.org/10.1186/s13063-023-07529-0.
4. Riva T, Engelhardt T, Basciani R, Bonfiglio R, Cools E, Fuchs A, Garcia-Marcinkiewicz AG, Greif R, Habre W, Huber M, Petre MA, von Ungern-Sternberg BS, Sommerfield D, Theiler L, Disma N; OPTIMISE Collaboration. Direct versus video laryngoscopy with standard blades for neonatal and infant tracheal intubation with supplemental oxygen: a multicentre, non-inferiority, randomised controlled trial. Lancet Child Adolesc Health. 2023;7:101–11. https://doi.org/10.1016/S2352-4642(22)00313-3.
5. McCann ME, de Graaff JC, Dorris L, Disma N, Withington D, Bell G, Grobler A, Stargatt R, Hunt RW, Sheppard SJ, Marmor J, Giribaldi G, Bellinger DC, Hartmann PL, Hardy P, Frawley G, Izzo F, von Ungern Sternberg BS, Lynn A, Wilton N, Mueller M, Polaner DM, Absalom AR, Szmuk P, Morton N, Berde C, Soriano S, Davidson AJ; GAS Consortium. Neurodevelopmental outcome at 5 years of age after general anaesthesia or awake-regional anaesthesia in infancy (GAS): an international, multicentre, randomised, controlled equivalence trial. Lancet. 2019;393(10172):664–77.
6. Cho E, Kwak JH, Huh J, Kang I-S, Ryu K-H, Lee SH, Ahn JH, Choi H-K, Song J. A comparative study using gastric ultrasound to evaluate the safety of shortening the fasting time before pediatric echocardiography: a randomized controlled non-inferiority study. J Anesth. 2024. https://doi.org/10.1007/s00540-024-03360-2.
7. Bouman AC, Ten Cate-Hoek AJ, Ramaekers BLT, Joore MA. Sample size estimation for non-inferiority trials: frequentist approach versus decision theory approach. PLoS ONE. 2015. https://doi.org/10.1371/JOURNAL.PONE.0130531.
8. Spencer AO, Walker AM, Yeung AK, Lardner DR, Yee K, Mulvey JM, Perlas A. Ultrasound assessment of gastric volume in the fasted pediatric patient undergoing upper gastrointestinal endoscopy: development of a predictive model using endoscopically suctioned volumes. Paediatr Anaesth. 2015;25:301–8. https://doi.org/10.1111/pan.12581.
9. Hayes J, Zuercher M, Gai N, Chowdhury AR, Aoyama K. The fragility Index of randomized controlled trials in pediatric anesthesiology. Can J Anesth/J Can D'anesthésie. 2023. https://doi.org/10.1007/s12630.
10. Ho AM, Sehgal A, Leitch J, Saha T, Mizubuti GB. Comment on: the fragility index of randomized controlled trials in pediatric anesthesiology. Can J Anesth. 2024;71(1):163–4.
11. Hayes J, Aoyama K. In reply: comment on: the fragility index of randomized controlled trials in pediatric anesthesiology. Can J Anesth. 2024;71:165–6.
12. Flight L, Julious SA. Practical guide to sample size calculations: an introduction. Pharm Stat. 2016;15:68–74. https://doi.org/10.1002/PST.1709.
13. Flight L, Julious SA. Practical guide to sample size calculations: superiority trials. Pharm Stat. 2016;15:75–9. https://doi.org/10.1002/PST.1718.
14. Flight L, Julious SA. Practical guide to sample size calculations: non-inferiority and equivalence trials. Pharm Stat. 2016;15:80–9. https://doi.org/10.1002/PST.1716.
15. Julious SA. Sample Sizes for Clinical Trials. 2nd ed. Boca Raton: Chapman and Hall/CRC; 2023.