Optimizing the study design of clinical trials to identify the efficacy of artificial intelligence tools in clinical practices

Qian Zhou [a,*], Yi-heng Cao [b], Zhi-hang Chen [c]

[a] *Department of Medical Statistics, Clinical Trials Unit, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Er Road, Guangzhou,510080, China*
[b] *ESIEE Paris, Cité Descartes BP99, Noisy-le-Grand 93162, France*
[c] *Department of Liver Surgery, The First Affiliated Hospital of Sun Yat-sen University, 58 Zhong Shan Er Road, Guangzhou, 510080, China*

A R T I C L E   I N F O

To the Editor,

We read with interest the recent article by Lin et al. [1] aimed to compare the diagnostic efficacy between CC-Cruiser a developed artificial intelligence (AI) platform [2] and ophthalmologists in real-world clinical settings in a randomized controlled trial (RCT). They concluded that CC-Cruiser had the capacity to assist doctors in clinical practice even though it showed less accuracy compared to senior consultants in diagnosing and making treatment decisions. However, we believe that the study design of the research could be improved by deeply considering the properness of using RCT.

The authors set a higher diagnostic accuracy for senior consultants than AI in the initial study design of sample size estimation, which meant they believed the performance of AI was not superior to that of senior experts. Even if the results of the study were positive, it would not be able to confirm the efficacy of AI and put it into "market." We recommend adopting a non-inferior design to confirm that AI is not inferior to seniors by setting an acceptable non-inferiority margin say 5% of diagnostic accuracy [3]. In the case where this condition is met and taking account of AI's advantages such as faster decision-making and lower-cost than seniors, if the non-inferiority was demonstrated, AI can be invested in the clinical settings where senior consultants are lacking. Further reading of the article identified an additional issue about study design that require attention. They reported that experts providing diagnosis were blinded to the group assignments to help prevent ascertainment bias. However, the study showed that the diagnostic accuracy from these experts were 4.1% higher than the estimated

95%. There might be a trial effect, that is, experts participated in this trial might be under pressure to "lose" against an AI and they would do their best to reflect their own level, which may be performed better than normal clinical practices. In this case, single arm diagnostic accuracy testing trial design would be more proper to identify the efficacy of AI by avoiding recruiting clinicians as a competing group.

Since the trend of including AI into clinical practices is massively increasing [4] and because Lin's study is one of the first published RCTs comparing the diagnostic efficacy of AI against experts, it will certainly have a far-reaching impact on future studies about AI tools. Well-designed phase 3 clinical trial is considered to be the final step for a drug or treatment being used in clinical practices. However, an inappropriate study design may prevent AI tools from further implementing in clinical practice. Therefore, we call for the improvement of study design and reporting guidelines of research about AI model implemented in clinical settings as soon as possible.

AI: Artificial Intelligence.
RCT: Randomized Controlled Trial.
CC-Cruiser: Congenital cataracts-Cruiser.

Ethics approval and consent to participate

Not applicable.

Consent for publication

Not applicable.

* Corresponding author.
  E-mail address: zhouq49@mail.sysu.edu.cn (Q. Zhou).

Availability of data and material

Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.

Funding

Not applicable.

Declaration of Competing Interest

None (all authors).

CRediT authorship contribution statement

Qian Zhou: Conceptualization, Methodology, Project administration, Supervision, Writing - original draft. **Yi-heng Cao:** Methodology, Project administration, Writing - review & editing. **Zhi-hang Chen:** Methodology, Project administration, Writing - review & editing.

Acknowledgements

References

[1] Lin H, Li R, Liu Z, et al. Diagnostic efficacy and therapeutic decision-making capacity of an artificial intelligence platform for childhood cataracts in eye clinics: a multicentre randomised controlled trial. E Clin Med 2019;9:55–63.

[2] Long E, Lin H, Liu Z, et al. An artificial intelligence platform for the multihospital collaborative management of congenital cataracts. Nat Biomed Eng 2017;1(2):0024.

[3] Piantadosi S. Clinical trials: a methodologic perspective. Wiley; 2006. 2nd ed..

[4] Sollini M, Antunovic L, Chiti A, Kirienko M. Towards clinical application of image mining: a systematic review on artificial intelligence and radiomics. Eur J Nucl Med Mol Imag. 2019. doi:10.1007/s00259-019-04372-x.