Special Issue

How Artificial Intelligence Can Transform Randomized Controlled Trials

Cecilia S. Lee[1] and Aaron Y. Lee[1,2]

[1] Department of Ophthalmology, University of Washington School of Medicine, Seattle, WA, USA
[2] eScience Institute, University of Washington, Seattle, WA, USA

Correspondence: Aaron Y. Lee, Department of Ophthalmology, University of Washington, Box 359607, 325 Ninth Avenue, Seattle, WA 98104, USA. e-mail: leeay@uw.edu

With the advent of deep learning (DL), the application of artificial intelligence (AI) and big data in healthcare has started transforming the way we approach medicine including clinical trials.[1,2] The randomized controlled trial (RCT) has been traditionally accepted as the most robust method of assessing the risks and benefits of any intervention.[3] However, the undertaking of an RCT is not always feasible due to the rarity of the disease, or time and costs that would impinge on the healthcare system.

AI is an academic discipline founded in 1956.[4] Machine learning (ML) is a subfield of AI that can learn complex relationships or patterns from data and make accurate decisions.[5] DL or deep artificial networks are a relatively new subfield of ML that takes advantage of powerful computational processing capacity provided by Graphic Processing Units and exponentially increasing datasets from medical records, images, multi-omics, and other "Big Data".[6] By feeding an enormous amount of data in training, a DL algorithm allows the model to alter its internal parameters between each neuronal layer to increase its performance. Applications of AI, DL in particular, have been successful in ophthalmic imaging research,[7–10] and the application of AI in RCTs may become reality in the near future.

Common pitfalls of unsuccessful RCTs include poor patient selection, inadequate randomization with residual confounders, insufficient sample size, and poor selection of end points.[11] With well-curated large datasets that incorporate clinical and multimodal imaging, AI models can be trained to select the potential study participants without relying on costly manual review to predict the natural history of each study participants with advanced statistical methods, and to assess study end points in a data-driven method. Given

these advantages, the application of AI has potentials for more efficient execution and greater statistical power than what would be expected from traditional RCTs.

First, ML models can drastically improve the patient selection process, thus lowering the burden of individual screening and need for large sample sizes. Recruiting the patients who meet precise selection criteria is crucial to avoid potential confounders or misclassifications. ML can combine multimodal data, such as imaging, laboratory, and other complex -omics data, to screen and select patients who match complex inclusion criteria, which can improve the recruitment efficiency. This is one of the areas in which the American Academy of Ophthalmology's Intelligent Research in Sight (IRIS) data will be utilized for RCT recruitment (personal communication, Flora Lum, MD).

In addition to the efficient selection process, having a sufficient sample size to enable detection of statistically significant differences between groups is critical. Many RCTs require a large sample size because the effect of the treatment in question is small.[12] AI has the potential in selecting "the ideal" patients for RCTs, who are "fast progressors" of the disease based on the AI's predictive algorithm. Thus, the expected effect size will be large and required sample size will be small resulting in a much shorter duration of RCTs. Selecting the "fast progressors" alone will limit the generalizability of the trial results; however, it may expedite the development of novel therapies, in particular for rare diseases.

Second, AI-generated end points have the potential to minimize measurement errors and analyze the data without human-imposed biases. Furthermore, algorithms may enable more sensitive quantification of key study end points than how they are traditionally measured. For example, central macular

thickness from optical coherence tomography (OCT) has been an important outcome in many RCTs but its reproducibility and correlation has been shown to vary among different methods of measurements (e.g. central subfield mean thickness vs. center point thickness) and different OCT devices.[13,14] More importantly, no standard method of quantifying paracentral or extrafoveal macular edema exists, even though this is an important end point in many noncenters involving retinal diseases, such as macular telangiectasia or branch retinal vein occlusions. Many studies have performed manual measurements of retinal thickness in one or several OCT slides, which may limit overall analyses. In contrast, an AI-generated algorithm has been shown to quantify the total amount of intraretinal cysts from entire OCT volumes in a fully automated fashion.[8]

Furthermore, AI models could generate new functional end points using structural data (e.g. OCT angiography from OCT and microperimetry from OCT) unlocking the potential of already archived data.[9,15] To illustrate, microperimetry (MAIA; Centervue) requires substantial test time and patient cooperation, thus limiting its use in clinical trials. In addition, microperimetry only tests 10° diameter area with 37 sensitivity points in the macula, which may not be sensitive or specific enough for clinical questions. By registering structural OCT and microperimetry test points together, AI was shown to predict microperimetry results from structural OCT with a mean absolute error of 3.36 dB.[9] More notably, DL models were able to generate continuous microperimetry predictions throughout the macula using structural OCT images. Therefore, applications of algorithms that can predict functional endpoints from structural end points may result in increased speed of evaluations and quality and/or quantity of end points.

Third, AI algorithms have the potential to enable direct measurement of treatment effects by taking a data-driven approach. Rather than expert-derived imaging markers being manually extracted from the imaging data, ML models could be applied directly to the outcome imaging data. Akin to Monte-Carlo permutation methods,[16] the labels for the clinical imaging from the control and treatment arms could be randomly shuffled, and ML models would attempt to be trained to predict whether the images came from the treatment or the control arm. If, in the unshuffled, original state, an ML model can be trained to accurately predict whether the images came from the treatment or the control arm statistically above the random shuffled states, then AI algorithms could directly measure a treatment effect in a data-driven fashion.

Finally, AI may allow the use of a synthetic control arm in the future. A frequent challenge in RCT is a sufficient enrollment of patients who meet the inclusion and exclusion criteria. Randomization is an essential aspect of a clinical trial in which a significant portion of participants are assigned to a control group. With sufficient data to train AI models to predict the natural history of each participant, substitution of the control arm by virtual controls may be possible cutting the recruitment goal by a significant amount. As a proof of concept, DL models have been able to predict what Humphrey visual field (HVF) would appear in up to 5.5 years from a single baseline HVF while ingesting clinical metadata.[17] The algorithm will need to be validated in independent populations but provides preliminary data that AI models could be used to predict disease progressions in synthetic controls. The prediction models would ingest the sum total of clinical, genetic, and imaging data to generate future progression of disease for each trial participant. Because of the data-driven method, unknown confounders may still be present among groups similar to RCTs. As an added benefit, this may increase the participation rate of the subjects who are reluctant about participating in trials because of the possibility of being in the placebo arm. The first step would be to incorporate a synthetic control arm without replacing the traditional placebo arm, so that the prediction of the synthetic arm could be evaluated prospectively without affecting the results of the RCT. In addition, a careful design and evaluation method for the prediction arm will be necessary to ensure that the same mistakes made of historical control arms from retrospective data are not recapitulated.

To allow fast development of AI algorithms, large-scale collaboration will be the first step to enable the storage of well-curated datasets, such as previous RCT data including multimodal imaging. Partnerships with pharma and imaging companies will accelerate this process. Development of standard minimum imaging protocols or testing methods with different manufacturers will be critical so routine clinical data can be used for novel research questions. Small steps, such as creating an "AI arm" alongside the usual study and control arms in future RCTs, to explore the potential will help validate the approach even when the main trial itself fails to meet the primary end point.

Many limitations still exist with this class of ML algorithms. The quality of algorithms is heavily dependent on the availability of large, well-labeled data, which may not be free from measurement error. The algorithms that bypass manual labels will be important by using more objective training targets. The "black-box" nature is another limitation. Methods

that explore the source of an AI decision tree, such as class activation maps[18] and occlusion testing,[7,19] will be key in integrating AI into RCTs.

Novel technologies have not been easily adopted in medicine traditionally, and AI will not replace RCTs. However, AI has the potential to improve and complement RCTs significantly in the future. The synergy among clinicians, researchers, and industries in collaborative efforts to share and collect standardized data, and allow AI algorithms to play major roles in RCTs may require paradigm shifts. However, these efforts will expedite the development of AI in ophthalmology, which will ultimately increase the quality of care that we provide for our individual patients.

Acknowledgments

Citation: Lee CS, Lee AY. How artificial intelligence can transform randomized controlled trials. Trans Vis Sci Tech. 2020;9(2):9, https://doi.org/10.1167/tvst.9.2.9

References

1. Luo J, Wu M, Gopukumar D, Zhao Y. Big data application in biomedical research and health care: a literature review. *Biomed Inform Insights*. 2016;8:1–10.
2. Yu K-H, Beam AL, Kohane IS. Artificial intelligence in healthcare. *Nat Biomed Eng*. 2018;2:719–731.
3. Chew EY. The value of randomized clinical trials in ophthalmology. *Am J Ophthalmol*. 2011;151:575–578.
4. Moor J. The Dartmouth College Artificial Intelligence Conference: the next fifty years. *AI Magazine*. 2006;27:87–87.
5. Wang S, Summers RM. Machine learning and radiology. *Med Image Anal*. 2012;16:933–951.
6. Jones LD, Golan D, Hanna SA, Ramachandran M. Artificial intelligence, machine learning and the evolution of healthcare. *Bone Joint Res*. 2018;7:223–225.
7. Lee CS, Baughman DM, Lee AY. Deep learning is effective for the classification of OCT images of normal versus age-related macular degeneration. *Ophthalmol Retina*. 2017;1:322–327.
8. Lee CS, Tyring AJ, Deruyter NP, Wu Y, Rokem A, Lee AY. Deep-learning based, automated segmentation of macular edema in optical coherence tomography. *Biomed Opt Express*. 2017;8:3440–3448.
9. Kihara Y, Heeren TFC, Lee CS, et al. Estimating retinal sensitivity using optical coherence tomography with deep-learning algorithms in macular telangiectasia type 2. *JAMA Netw Open*. 2019;2:e188029.
10. Ting DSW, Cheung CY-L, Lim G, et al. Development and validation of a deep learning system for diabetic retinopathy and related eye diseases using retinal images from multiethnic populations with diabetes. *JAMA*. 2017;318:2211–2223.
11. Nichol AD, Bailey M, Cooper DJ. Challenging issues in randomised controlled trials. *Injury*. 2010;41:S20–S23.
12. Stolberg HO, Norman G, Trop I. Randomized controlled trials. *AJR Am J Roentgenol*. 2004;183:1539–1544.
13. Wells JA, Glassman AR, Ayala AR, et al. Aflibercept, bevacizumab, or ranibizumab for diabetic macular edema: two-year results from a comparative effectiveness randomized clinical trial. *Ophthalmology*. 2016;123:1351–1359.
14. Browning DJ, Glassman AR, Aiello LP, et al. Optical coherence tomography measurements and analysis methods in optical coherence tomography studies of diabetic macular edema. *Ophthalmology*. 2008;115:1366–1371.e1.
15. Lee CS, Tyring AJ, Wu Y, et al. Generating retinal flow maps from structural optical coherence tomography with artificial intelligence. *Sci Rep*. 2019;9:5694.
16. Tusher VG, Tibshirani R, Chu G. Significance analysis of microarrays applied to the ionizing radiation response. *Proc Natl Acad Sci USA*. 2001;98:5116–5121.

17. Wen JC, Lee CS, Keane PA, et al. Forecasting future Humphrey visual fields using deep learning. *PLoS One*. 2019;14:e0214875.

18. Zhou B, Khosla A, Lapedriza A, Oliva A, Torralba A. Learning deep features for discriminative localization. In: *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*. 2016:2921–2929.

19. Zeiler MD, Fergus R. Visualizing and Understanding Convolutional Networks. In: *Computer Vision – ECCV 2014*. Springer International Publishing; 2014:818–833.