ISPOR Poster Code: RWD145

Machine Learning and Artificial Intelligence for Clinical Trial Optimization: A Review of Opportunities to Leverage Real World Data

Christina Mack[1], Jimeng Sun[2], Zifeng Wang[2], Chufan Gao[2], Kathryn Rough[1], Lucas Glass[3]

[1]IQVIA Center for Advanced Evidence Generation; [2]University of Illinois, Urbana-Champaign; [3]IQVIA Analytics Center of Excellence

Background & Objective

- Clinical trials for novel therapeutics are expensive, time-consuming, and have an end-to-end success rate of under 8% [1].
- Applications of artificial intelligence and machine learning (AI/ML) to trial design and conduct have the potential to increase trial efficiency, and the use of real world data (RWD) is often critical to the development of these novel applications.

Objective

The objective of this study was to comprehensively review applications of AI/ML that used RWD to improve clinical trial design and conduct.

Methods

- We reviewed published work on AI/ML-based technologies for clinical trials, focusing on applications trained using RWD.
  - › We survey the field, with concrete examples of applications.
- Open challenges and opportunities for the field are summarized.

Results

- AI/ML can be used throughout the drug life cycle (Figure 1); for clinical trials, RWD from electronic health records is particularly essential.
- After reviewing 65 published examples, we identified four overarching domains of applications:
  - › Improved participant selection & recruitment
  - › Enhanced trial management, including automation of data management, processing, and harmonization
  - › Enabling remote patient monitoring (e.g., processing wearables data)
  - › Synthetic patient generation & digital twins (i.e., *in silico* trials) [Figure 2]
- Challenges:
  - › Limited availability and aggregation of RWD
  - › Limited model generalizability/transportability
  - › Potential for fairness issues and bias
  - › Lack of familiarity with AI/ML methods among trialists
  - › Limited prospective evaluations of novel designs
- Opportunities for advancing RWD-based AI/ML for clinical trials:
  - › Transfer learning to accommodate limited disease- or domain-specific data
  - › Increased use of multi-modal data in AI/ML model training
  - › Applications of federated learning when data is distributed (e.g., for patient privacy or proprietary reasons)

Figure 1. AI/ML throughout the drug lifecycle: four domains of identified clinical trial applications

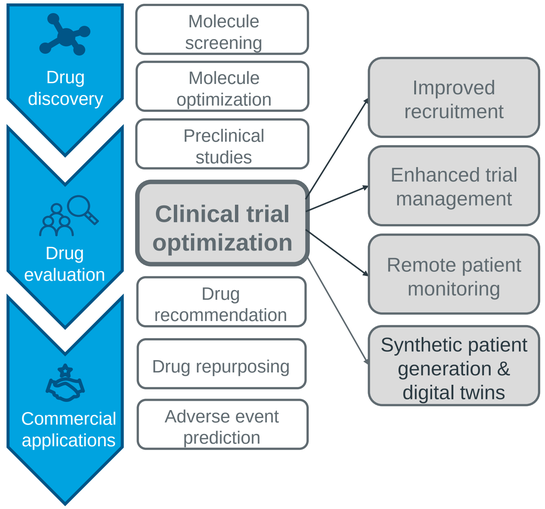

Figure 2. AI/ML approaches for leveraging RWD to generate synthesize trial-relevant data

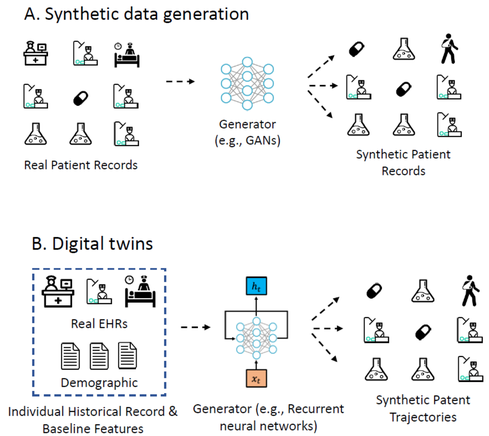

Table 1. Example AI/ML applications proposed in each domain identified in the review

Topic	Application	Task description	AI/ML methodology
Improved patient selection & recruitment	Rajpurkar et al [2]	Anticipate antidepressant treatment response based on pre-treatment symptoms and electroencephalography	Gradient-boosted decision trees
Enhanced trial management	Yu et al [3]	Extraction of free-text clinical trial eligibility criteria into structured query language (SQL)	Long short term memory (LSTM) model
Enabling remote patient monitoring	Tucker et al [4]	Identification of treatment non-adherence in Parkinson's disease patients using motion sensing device	Gradient-boosted decision trees
Synthetic patient generation & digital twins	Walsh et al [5]	Model progression of multiple sclerosis to generate a synthetic placebo-controlled trial arm	Conditional restricted Boltzmann machine (CRBM) model

Conclusions

- AI/ML technologies trained with RWD have the potential to reduce costs and optimize clinical trials through increased study efficiency.
  - › AI/ML can improve multiple aspects of trial delivery, including participant selection & recruitment, trial management, remote patient monitoring, and synthetic patient generation.
- Most research reviewed was in the proof-of-concept or feasibility stage with relatively few documented applications in active clinical trial delivery.
  - › As a result, there is currently limited quantification of the benefits these technologies have on clinical trial conduct and further research is warranted.

References

[1] Biotechnology Innovation Agency. Clinical development success rates and contributing factors 2011-2020 (2021). Available from: https://go.bio.org/rs/490-EHZ-999/images/ClinicalDevelopmentSuccessRates2011_2020.pdf

[2] Rajpurkar P, Yang J, Dass N, Vale V, Keller AS, Irvin J, Taylor Z, Basu S, Ng A, Williams LM. Evaluation of a Machine Learning Model Based on Pretreatment Symptoms and Electroencephalographic Features to Predict Outcomes of Antidepressant Treatment in Adults With Depression: A Prespecified Secondary Analysis of a Randomized Clinical Trial (2020). JAMA Network Open. 3(6):e206653. Available from: https://pubmed.ncbi.nlm.nih.gov/32568399/

[3] Yu X, Chen T, Yu Z, Li H, Yang Y, Jiang X, Jiang A. Dataset and Enhanced Model for Eligibility Criteria-to-SQL Semantic Parsing (2020). *Proceedings on the 12th Conference on Language Resources and Evaluation*. 5829-5837. Available from: https://aclanthology.org/2020.lrec-1.714.pdf

[4] Tucker CS, Behoora I, Nembhard HB, Lewis M, Sterling NW, Huang X. Machine learning classification of medication adherence in patients with movement disorders using non-wearable sensors (2015). *Computers in Biology and Medicine*. 66:120-34. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5729888/

[5] Walsh JR, Smith AM, Pouliot Y, Li-Bland D, Loukianov A, Fisher CK. Generating Digital Twins with Multiple Sclerosis Using Probabilistic Neural Networks (2020). [Preprint]. Available from: https://arxiv.org/abs/2002.02779