WHITE PAPER

cool clinical

Consortium for Clinical Science and AI

Machine Learning for Precision Medicine and Clinical Stage Drug Development

Felix Beacher, PhD
November 2020

About Cool Clinical

Cool Clinical is a non-profit consortium of clinical and computational scientists. Cool Clinical publishes original articles and reports for businesses, the public sector, NGOs and the general public. Our goal is to advance the conversation on AI applications to clinical science.

To learn more about the research of Cool Clinical, visit www.coolclinical.com

Contents

Machine Learning and Clinical Science

ML is a subset of artificial intelligence, and involves the development of computer algorithms that can be 'trained' on data. ML algorithms build complex mathematical models, based on relatively simple rules, to make predictions. ML can be successful in many areas where conventional computer programming is ineffective due to the complexity of the problem. A classic example of this is computer vision, where it is difficult to develop an 'expert system' to perform object recognition, e.g., distinguishing between cats and dogs. ML has been applied to many areas of technology, science and medicine and its application is projected to grow rapidly in coming years.

One area of medicine in which ML has proved highly successful is diagnostics. The problem of under-diagnosis is very significant for many diseases (Singh et al 2014). For example, underdiagnosis rates are estimated to range from 20% for dementia and cirrhosis, to 90% for depression and osteoarthritis (Falagas et al 2007). Further, for rare diseases, it takes an average of five years between the presentation of the first symptoms and a correct diagnosis, and a correct diagnosis may often never occur. ML has the potential to revolutionize diagnostics, especially where diagnostic procedures require special expertise and when they can be time consuming and/or expensive. ML can use patterns in data that is difficult to notice by the human eye. ML-based diagnostics, exceeding the performance of the best human clinicians, have been developed for a range of disorders, including neurodegenerative diseases (Myszczynska et al 2020), cardiovascular disease (Denaxas and Morley 2015) and various types of cancer (Savage 2020). It is in diagnostics that the FDA has approved ML-based clinical products, for example image classifiers based on MRI, CT, or X-ray data, such as mammography. Typically, these systems (devices and/or software) provide a provisional diagnosis to be checked by a human expert.

Machine Learning and Precision Medicine

'Precision medicine' refers to the tailoring of medical treatment depending on the individual characteristics of the patient, in a way which is objective and data-based. For any drug treatment, some patients will respond very well, others moderately well, and others will respond poorly. If these responses can be successfully predicted, the best treatment option could be selected. This is a basic implementation of precision medicine.

In general, precision medicine remains theoretical. Normal clinical practice may attempt to tailor treatment to some degree, but based on clinical judgement which suffers from subjectivity and limited personal experience. Often a 'one-drug-fits-all' approach is used, where a diagnosis leads to a certain type of treatment. Alternatively, a trial-and-error approach may be used, in which different treatment options and/or dosages are tried in a more-or-less unguided fashion.

ML has the potential to deliver a paradigm shift toward precision medicine (Mesko 2017, Krittanawong et al 2017). Currently, precision medicine is mostly associated with individualized treatment on the basis of subtypes of the patient's cancer and genetic mutations. For example, the poly ADP ribose polymerase (PARP) inhibitor olaparib ('Lynparza') is approved by the Food and Drug Administration (FDA) as a monotherapy for ovarian cancer in women with BRCA1/2 mutations, and for breast cancer in women with hereditary BRCA1 or BRCA2 mutations (Ledermann et al, 2014; Krzyszczyk et al 2019). Despite the current focus of precision medicine on oncogenes, ML-based precision medicine could use a wide range of data, including symptoms, imaging tests, biochemistry results and -omics (proteomics, transcriptomics, metabolomics, etc.; Williams 2015; Lu et al 2014).

In addition to the benefit of precision medicine for patients, the ability to predict treatment responses in clinical practice could also offer a competitive advantage for pharmaceutical companies. For example, if a drug is shown to be superior to its competitor products for a newly identified subpopulation (all other things being equal) patients within that subpopulation may be transferred from other treatments to it.

The Use of Machine Learning for Clinical Trials

The Costs of Drug Development

The ability to predict patient drug responses could improve the cost efficiency of clinical-stage drug development. This is an important issue because drug development has become extremely expensive in recent decades. Estimates of the average development cost of a new drug vary from $1.3 billion (Wouters et al, 2020) to $2.6 billion (in 2013 US dollars; DiMasia et al 2016). Not only are drug development costs very high, investigational drugs have a significant failure rate (Dowden and Munro 2019, Hay et al 2014; Wong et al 2019), especially in the field of oncology (Harrison 2016). These factors represent a severe threat to existing models of pharmaceutical development. It has even been claimed that we are entering a 'post-blockbuster era' (Harrer et al 2019), in which the costs of new drug development are becoming prohibitive, even for large drug companies.

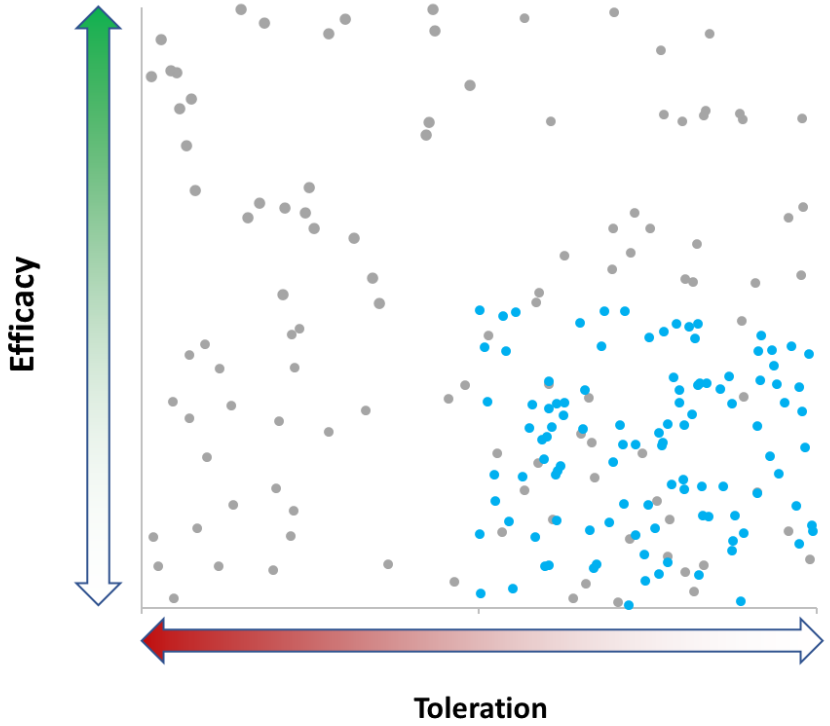
Clinical Trial Enrichment

ML-based systems which can predict Phase III clinical trial outcomes could help address the challenges of the costs and uncertainty surrounding drug development. The FDA defines 'clinical trial enrichment' as 'the prospective use of any patient characteristic to select a study population in which detection of a drug effect (if one is in fact present) is more likely than it would be in an unselected population'. In 2019 the FDA issued guidance to 'assist industry in developing enrichment strategies that can be used in clinical investigations intended to demonstrate the effectiveness of drug and biological products' (https://www.fda.gov/media/121320/download).

'Predictive enrichment' refers to a type of clinical trial enrichment, which only includes patients in clinical trials who are likely to positively respond to the drug treatment. 'Prognostic enrichment' is a variant, which only includes patients in clinical trials who are likely to have a particular disease-related event, e.g., a heart attack. ML can potentially be applied to both types of clinical trial enrichment.

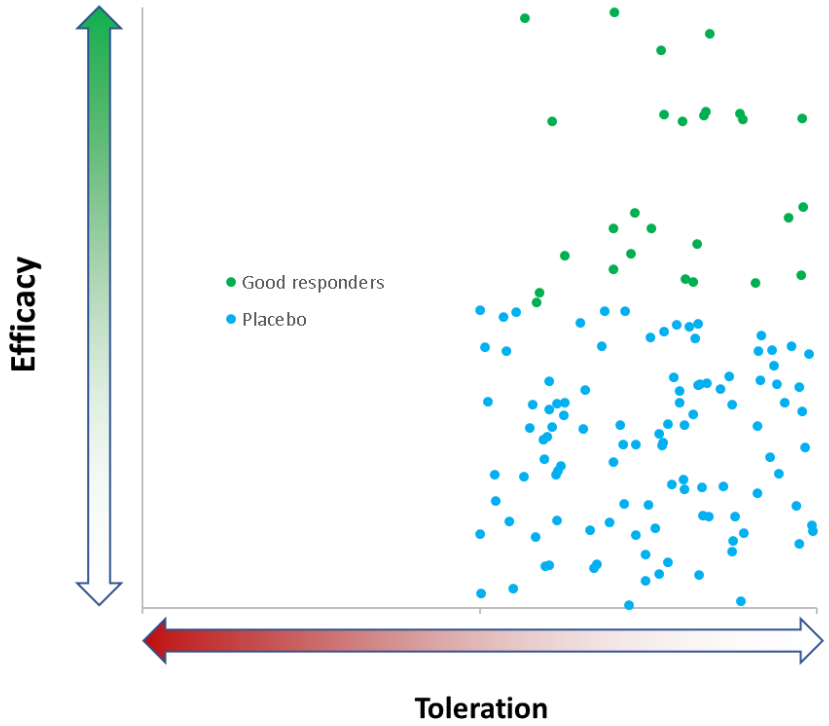
Figure 2 illustrates the relationship between drug safety and efficacy, with and without successful predictive enrichment. Figure 2A illustrates the results of a typical clinical trial, where a drug has greater efficacy than placebo, but worse safety. This poses a dilemma for regulators who have to balance the risks with the likely benefits of the drag compared to placebo. Figure 2B illustrates the results of a clinical trial, after successful enrichment, where the drug has greater efficacy than placebo, but a similar safety profile as placebo. This presents a clearer choice for regulators who have to balance the risks with the likely benefits of the drag compared to placebo.

A: Drug vs placebo with no patient selection



Drug is more efficacious than placebo, but less safe. Thus, regulators have to assess the balance of risk and benefit.

B: Drug vs placebo after successful patient selection ('predictive clinical trial enrichment')



The drug is more efficacious than placebo, but has a similar safety profile to placebo. Thus, regulators have to assess the balance of risk-benefit to approve a drug.

Figure 1 The relationship between drug safety and efficacy, without and with successful predictive enrichment

The FDA guidance identifies various regulatory issues related to enrichment strategies, for example:

1. The method for patient selection should be stated prior to study initiation, to avoid biasing the study results.
2. An enrichment strategy should allow the intended patients to be identified in the real world.
3. A sponsor should collect secondary data on marker-negative patients to assess whether they could also benefit from the drug

The FDA Guidance considers different strategies for predictive enrichment, including the following:

1. *Empirical strategies*, where the selection of likely responders is based on observed response during screening or prior experience with the drug or related drugs. There is no mechanistic understanding of the basis for the observed differences in drug response.

2. *Pathophysiologic strategies* where the selection of likely responders is based on an assessment of the patients' individual physiology/pathophysiology in the context of an understanding of the disease mechanism. Pathophysiological indicators can be biomarkers (including relevant gene mutations), imaging data, demographic features or clinical features.

3. *Empirical genomic* strategies where the selection of likely responders is based on genetic/genomic/proteomic patterns, but the mechanism linking these patterns to the disease is unknown, hence 'empirical'.

Predictive enrichment is potentially appropriate for both phase II and phase III drug trials, as both measure efficacy. Also, the data from phase II trials could potentially be used to support the predictive enrichment of phase III trials.

Clinical Trial Enrichment for Improving the Chance of Regulatory Approval

The main cause of regulatory failure is the inability to demonstrate efficacy (Figure 2; Harrison 2016). In some cases, efficacy of drugs may be real but fail to be demonstrated due to issues such as flawed study design, inappropriate endpoints, underpowered studies or a patient population with a high degree of heterogeneity (including patients poorly suited to the treatment).

Clinical trial enrichment, by identifying subpopulations in which investigational treatments efficacious and safe, could improve the chance that an investigational drug will achieve regulatory approval compared to a non-selected patient population. This strategy assumes that the same procedure for patient selection is used in the clinical trial as in post-approval clinical practice. On the same basis, investigational drugs previously rejected by regulators could be 'rebooted', by repeating a phase III clinical trial but within a predicted subpopulation.

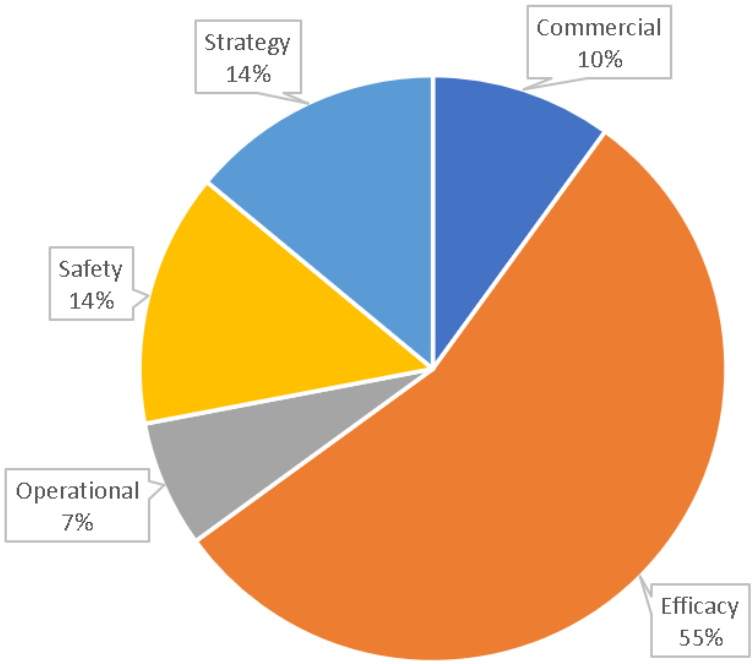


Figure 2 Reasons for Failure of Phase III Trials (data from Harrison 2016)

Clinical Trial Enrichment for Reducing Clinical Trial Sample Size

The largest part of the cost of drug development is represented by clinical trials, accounting for around half of the total cost. Phase III trials, in particular, are by far the most expensive type of clinical trial, due to their large size (usually in the thousands) and their longer-term nature, compared to early phase trials. By improving the safety and efficacy profile of a drug within a newly-identified subpopulation, clinical trial enrichment can mean that required sample sizes can be smaller, resulting in potentially large costs savings to drug companies (Simon and Maitournam 2004). Importantly this has been explicitly recognized by the FDA (2019 FDA Guidance on enrichment strategies).

In general, the higher the degree of population selection, and the smaller the treatment effect in the marker-negative population, the more the sample size of a clinical trial can be reduced.

Types of variables

The FDA guidance gives examples of the kinds of variables which could be used in the development of enrichment strategies:

- demographic variables such as sex and age
- disease characteristics
- presence of genetic markers
- physiological measurements
- concomitant illnesses
- biomarker measurements
- co-morbidities
- history of drug responses

Clinical Trial Enrichment and Machine Learning

Currently there is no FDA guidance that explicitly covers ML in relation to clinical trial enrichment. Also, the FDA has not yet approved drugs based on clinical trials where ML has been used to identify novel patient subpopulations. However, a 2020 FDA publication *Machine-Learning-Derived Enrichment Markers in Clinical Trials* encourages research in this area (https://isctm.org/public_access/Feb2020/Presentation/Millis-Presentation.pdf). This document states that '[there are] interesting questions about how the FDA would provide oversight for a drug development program in which a machine learning algorithm has a key role in identifying the target population for the drug.' The current lack of FDA approvals for ML-based trial enrichment is not to do with principle, but rather that there is 'Little experience in FDA with evaluating study protocols that use ML-based methods for enriching the study population in a clinical trial'. The document states that 'Sponsors considering the use of ML-based classifiers in drug development should seek consultation from the FDA during the earliest stage possible in the development program'.

Other Applications for Machine Learning in Clinical Trials

ML for Selecting 'Good Subjects' for Clinical Trials

In addition to prognostic enrichment strategies, ML could be applied in other ways to promote the efficiency of clinical trials. For example, ML could potentially be used to select patients who are more likely to adhere to the study procedures. This kind of strategy could be of particular importance for trials where study procedures are challenging. Another example is that ML could be used to identify potential participants who are likely to have highly variable blood biochemistry measurements over time (Banda et al 2018). In all such cases, it is essential that such selection methods are used *before randomization* to avoid bias.

ML for Clinical Trial Recruitment

Recruitment is a major challenge for phase III clinical trials, because they require large numbers of participants (typically over 1000) and the eligibility criteria are often highly restrictive, especially for rare diseases. The challenge is compounded by the fact that patient databases used for recruitment may record information in ways which are be unstructured and inconsistent. Consequently, clinical trial recruitment is time consuming, and may result is delays, with a significant financial impact. Further, clinical trials may suffer from under-recruitment, leading to insufficient statistical power which can lead to rejection by regulators.

ML systems can help address the challenge of patient recruitment in various ways (Fogel 2018), for example, the automated search of patient databases using ML techniques such as Natural Language Processing (NLP) and optical character recognition (OCR). Such techniques can potentially be developed into systems which can automatically mine clinical trial databases or even social media content, to identify eligible people who are searching for clinical trials to take part in. The real-world success of such systems has already been demonstrated (Helgeson et al 2018), although challenges remain on the development and scaling of such systems.

Summary

To summarize, the ability to predict outcomes of drug treatment has various potential benefits:

1. To implement precision medicine, i.e., provide treatment with increased chance of success
2. To increase the chance of gaining regulatory approval for a new drug, by improving efficacy and safety within a selected subpopulation
3. To lower the costs of clinical-stage drug development, by allowing smaller sample sizes.

ML offers the potential to develop accurate systems for making these predictions.

References

Banda et al 2018. Advances in Electronic Phenotyping: From Rule-Based Definitions to Machine Learning Models. Annual Review of Biomedical Data Science. Vol. 1:53-68

Denaxas and Morley 2015. Big biomedical data and cardiovascular disease research: opportunities and challenges. Eur. Heart. J. Qual. Care Clin. Outcomes 1, 9–16.

DiMasia et al 2016. Innovation in the pharmaceutical industry: New estimates of R&D costs. Journal of Health Economics, 47: 20-33

Dowden and Munro 2019. Trends in clinical success rates and therapeutic focus. Nature Reviews Drug Discovery 18, 495-496

Falagas et al 2007. Under-diagnosis of common chronic diseases: prevalence and impact on human health. Int J Clin Pract 61(9):1569-79.

Fogel 2018. Factors associated with clinical trials that fail and opportunities for improving the likelihood of success: a review. Contemp. Clin. Trials Commun., 11, pp. 156-164

Harrer et al 2019. Artificial Intelligence for Clinical Trial Design. Trends in Pharmacological Sciences. Special Issue: Rise of Machines in Medicine. Volume 40, Issue 8, August 2019, Pages 577-591

Harrison 2016. Phase II and phase III failures: 2013–2015. Nature Reviews Drug Discovery 15, 817–818

Hay et al 2014. Clinical development success rates for investigational drugs. Nature Biotechnology 32, 40-51

Helgeson, et al 2018. Clinical performance pilot using cognitive computing for clinical trial matching at Mayo Clinic. J. Clin. Oncol., 36, Article e18598

Krittanawong et al 2017. Artificial Intelligence in Precision Cardiovascular Medicine. Journal of the American College of Cardiology. 69 (21): 2657–2664.

Krzyszczyk et al 2019. The growing role of precision and personalized medicine for cancer treatment. Technology (Singap World Sci). 6(3-4): 79–100.

Ledermann et al. Olaparib maintenance therapy in patients with platinum-sensitive relapsed serous ovarian cancer: A preplanned retrospective analysis of outcomes by BRCA status in a randomised phase 2 trial. Lancet Oncol 15, 852–861 (2014).

Lu et al 2014. Personalized medicine and human genetic diversity. Cold Spring Harbor Perspectives in Medicine, 4 (9).

Mesko 2017. Expert Review of Precision Medicine and Drug Development. Journal Expert Review of Precision Medicine and Drug Development. 2 (5): 239–241.

Myszczynska et al 2020. Applications of machine learning to diagnosis and treatment of neurodegenerative diseases. Nature Reviews Neurology volume 16, pages440–456

Savage 2020. Another set of eyes for cancer diagnostics. Nature, Vol 579, S15.

Simon and Maitournam, 2004. Evaluating the Efficiency of Targeted Designs for Randomized Clinical Trials, Clin Cancer Res, 10:6759–6763.

Singh et al 2014. The frequency of diagnostic errors in outpatient care: estimations from three large observational studies involving us adult populations. BMJ Qual. Saf. 23, 727–731

Williams 2015. Capturing cancer's complexity. Proc Natl Acad Sci; 112(15):4509-11.

Wong et al. Estimation of clinical trial success rates and related parameters. Biostatistics. 2019; 20: 273-286.

Wouters et al, 2020. "Estimated Research and Development Investment Needed to Bring a New Medicine to Market, 2009-2018". JAMA. 323 (9): 844–853.

About the author

Felix Beacher | felix@coolclinical.com

Felix Beacher is the founder of Cool Clinical. Felix has a PhD in neuroscience and has worked in drug development for neurodevelopmental and neurodegenerative disorders and cancer.

Legal disclaimer

This publication has been written in general terms and is not intended to be relied on to cover specific situations. Any application of the information given in this publication will depend upon the particular circumstances involved. As such, we recommend that professional advice is sought before acting or refraining from acting on any of the contents of this publication. This publication and the information contained herein is provided “as is”. Cool Clinical makes no express or implied warranties that this publication is error-free or meet any particular criterion of performance or quality.

Cool Clinical accepts no duty of care or liability for any loss to any person acting or refraining from action as a result of any material in this publication.

