AI in Drug Discovery and Development

Prof. Rajnish Kumar
Dept. of Pharmaceutical Engineering and Technology
IIT-(BHU), Varanasi

Week-08
Lecture-37

Welcome to the course on AI In Drug, Discovery, and Development. In today's session, we will talk about patient recruitment, certification, and retention, so by the end of this lecture, you will be able to understand the end-to-end patient journey. And the importance of patient-centric approaches in clinical trials: learn strategies to improve patient identification and streamline recruitment. Explore how AI enables accurate patient stratification for enhanced trial outcomes. Examine real-world use cases showcasing AI's impact on clinical trial processes, as well as identify challenges in AI adoption within clinical trials. So, let us take a look at some of the key terminologies.

The first one is patient identification. So, it is the process of systematically locating individuals who meet predefined inclusion and exclusion criteria for a clinical trial, often utilizing electronic health records, registries, or predictive algorithms. Patient recruitment is another term that refers to the act of enrolling eligible participants in a clinical trial through targeted outreach, communication, and engagement strategies, ensuring compliance with ethical and regulatory standards. So, these two steps are quite important.

As I said, like we discussed in an earlier session as well, whenever we are conducting a clinical trial, we need to identify the participants in that clinical trial. So those participants might need to have some specific disease conditions or they need to fulfill some criteria, actually. So, another term is "patient stratification." So, it is a methodological approach to classify patients into distinct subgroups based on clinical, genetic, demographic, or biomarker data to ensure balanced trial arms and improve the precision and efficacy of interventions. And then we have another term, which is patient retention.

So this is the maintenance of participants' involvement throughout the duration of a clinical trial, which is achieved through strategies that minimize dropout rates and ensure adherence to the study protocol. And then there is another key term called informed consent, which is a fundamental ethical and legal requirement in clinical research. wherein participants are provided with comprehensive information about studies, objectives, procedures, potential risks and benefits, and their rights, enabling voluntary and informed participation. So, informed consent is also a very important aspect of, you know, an important part of a clinical study. Where the patients are being apprised of all the information related to the study, and then only will they voluntarily take part in that study.

So, if you look at the journey of a patient or volunteer in a clinical trial, it starts with, you know, the recruitment. So that can be processed through social media as well. The companies advertise that we are looking for participants in this study, and those should be, for example, aged between this and this. They should have this condition and all the requirements they set. And then they encourage people to participate in it.

So once this advertisement is communicated to the public, the people who have shown interest are actually screened. So, they might be done with online questionnaires, web-based portals, or through video conferencing for the screening, and after that, those who meet the criteria for enrollment. So, they are being enrolled. And then there can be online informed consents through web-based portals or through video conferencing as well. And then the patients are being educated about the clinical trial study.

That can also happen through the web-based portals or videos. And then the initial visit to the center, to the clinical trial site, can also happen through virtual reality as well. After this initial visit, the intervention begins and then treatment starts. And then the data is being collected so that it can also be collected through the online surveys, you know, wearable sensors, or the medications. And then, after the data is collected, the study visit, which involves visiting the clinic for all those measurements of the endpoints, will take place.

And then this continuously goes on until the study is completed or there is any adverse event leading to the closure of that study. And those adverse events, those patients can report through the web portals, and this is also, you know, one of the things which is also known as pharmacovigilance, actually. This is the journey of the patient or a volunteer in the whole clinical trial. So, what are some of these things? These are some of the challenges that hinder patient recruitment in the clinical trial. So, more than 50 percent, around 55 percent of the participants, do not meet the inclusion criteria.

Because most of the time those studies need someone with, you know, predefined criteria. So, they do not meet that. So, that is one of the major problems, a major challenge that makes it difficult for a clinical trial to recruit participants. And then another major issue is the decline in the participants. So, people usually do not participate.

So, they say that we do not want to participate because this might happen due to their lack of knowledge about the importance of clinical trials. Or they have some kind of, you know, some fear in their mind. Some other challenges are also present, such as the non-compliance with the protocol. And then, not being willing to stay in the locality is another challenge; it is actually another region. And then they can get recruited in another study sometimes; you know, they can get recruited in another study as well.

Especially when they are, you know, having those pathologies that are actually in demand for clinical trials. And then sometimes they meet the exclusion criteria during the study as well, so those can be excluded during the study. And then they have poor performance status as well. So, around 3 percent of the participants stopped participating in the trial because of poor performance, and there are some other reasons as well. So, you can see that recruiting a patient in a clinical trial is actually quite a big challenge.

So then, there are ways out for patient recruitment, and some of them include awareness and education. We need to increase public understanding and dispel myths about clinical trials through awareness campaigns and healthcare partnerships. The eligibility criteria require us to balance strict safety standards with flexible designs to widen the pool of eligible participants. So, we are keeping the eligibility criteria very narrow. So, then, of course, we will not find a lot of people who can participate in the study.

So, if we are keeping this eligibility criteria a little bit flexible so that we can get, you know, enough number of participants. And then with accessibility, we can overcome the geographic and logistic barriers using decentralized trial models. And the travel support, which we have seen in COVID-19 cases, is that decentralized clinical trials are the future as well. And then diverse populations boost representation by engaging underserved communities with culturally sensitive outreach. And then motivation, where we can highlight benefits like innovative treatment and incentives to tap into participants' intrinsic motivators.

So, these are, you know, some of the ways by which patient recruitment can be enhanced. Okay, another thing is stratification. So why does stratification matter in clinical trials? Because we need to have greater precision, you know. So, it reduces variability by grouping similar patients, boosting the statistical power and clarity of treatment effects. And by the way, stratification is when we have, for example, 1,000 participants in a clinical trial.

So we actually stratify them into some groups based on several factors, which we will discuss in the following slides. So, then targeted insights, we need to identify which subgroups benefit most, enabling more personalized. Effective therapies, smarter trial design, which helps define inclusion and exclusion criteria, support adaptive design and minimize confounding factors. And then better safety and efficacy clarify how different groups respond, improving the detection of adverse effects, therapeutic benefits, and regulatory and market advantages. It supports regulatory approvals and enables targeted product positioning for specific patient populations.

So these are, you know, different kinds of patient stratification approaches. So, these can

be stratified based on the clinical manifestations. It is called clinical stratification. They can be stratified based on demographics. So, demographic stratification, genetic stratification, or genomic stratification, where those patients are grouped into groups based on their genetic makeup.

And biomarker-based stratification based on the biomarkers implemented in the study, we can classify them into groups for behavioral or psychosocial stratification and digital phenotyping. or AI-driven stratification, which is, of course, nowadays a very popular thing. So, talking about clinical stratification, in this case, patients are grouped based on their clinical characteristics. or disease-related parameters in order to balance trial arms for comparative disease burden, assess treatment efficacy across different severity levels, and reduce variability in clinical outcomes. And the criteria for certification we use generally include, for example, the disease stage, as in the case of cancer.

So at what stage of cancer are they? So, based on stages 1 through 4, they can be certified or based on the severity of the disease. Like in the case of asthma, whether they have mild, moderate, or severe asthma depends on the basis of comorbidities, such as whether they have diabetes, hypertension, or any other disease. or any other disease, along with the disease that is, you know, the main criterion for inclusion in the clinical trial. and the duration of the illness, like how long they have been, you know, sick with this disease. So, an example can be in a heart failure trial where patients may be classified by New York Heart Association functional class 1 to 4 to evaluate treatment response across different severities.

And then you have demographic stratification. So, this stratification is based on the demographic factors that may influence disease progression. For treatment response and in order to ensure demographic diversity to detect subgroup-specific effects or disparities, or to meet regulatory requirements like the FDA or ICH guidelines. So, the criteria for this stratification can be the age group. Whether they are in pediatric age, adult, or geriatric, sex or gender can also be one of the criteria for demographic stratification, along with ethnicity or race and geographic or socioeconomic background.

So, an example of demographic stratification is that a vaccine trial may stratify by age groups to ensure safety and efficacy data are generated for both younger and older populations. And then the next one is the genetic or genomic stratification, where the patients are grouped based on genetic variants and gene expression profiles. A genomic signature in order to identify patients most likely to benefit from the targeted therapies, minimize adverse drug reactions, and enable personalized treatment. So, some of the criteria can be single nucleotide polymorphisms, whether some people have the presence of the SNPs in their genomic gene signature or not. Based on that, they can be, you know,

stratified.

Germline or somatic mutations, such as BRCA1 or BRCA2 or EGFR mutations, and gene expression profiles, like Oncotype DX or the pharmacogenomic markers. And then you have the biomarker-based certification where the certification is based on measuring biomarkers that reflect disease state, drug response, or prognosis in order to identify likely responders and non-responders, monitor treatment efficacy, and guide dose or regimen selection. So, some of the criteria can include protein markers such as HER2 or PD-L1. Blood-based biomarkers like CRP or HbA1c, or imaging biomarkers like tumor burden in a PET scan, or metabolomics or proteomic signatures. So, an example of biomarker-based stratification is that breast cancer patients can be stratified based on HER2 expression to determine eligibility for trastuzumab (Herceptin) therapy.

So this can be one example of patient stratification using genetic and genomic biomarkers. So, the top panel shows the population-based stratification, which identifies genetic biomarkers to classify patients into distinct genetically defined groups with overlapping phenotypes. So, you can see that these are the population-based PGT approaches where the positive tests are responsive to the therapy. So, they will get standard therapy if they are, you know, not responding to the therapy. So, they will get a dose modification, and then we will get another dose for the same therapy based on the genomic biomarkers.

So, they can be in this one; actually, it integrates multi-omics data to define individual genomic signatures for targeted personalized therapy. So, you can make up based on the individual multi-omics you know. So, one can decide which person will get which therapy. So, this is like more, you know, targeted personalized therapy can be defined and decided using this approach. And then there is another kind of stratification that is behavioral or psychosocial, where stratification is based on behavioral traits.

or psychosocial factors that may influence treatment adherence or outcomes in order to tailor interventions and support mechanisms. Understand how known biological factors affect the outcomes and improve retention and compliance in the trial. So, some of the criteria for this behavioral or psychosocial stratification could be lifestyle factors like whether a person is smoking, using alcohol, or is physically active, or they can include mental health status as well, such as depression, anxiety, or socioeconomic status. Medication adherence behavior and social support systems. So, an example of this behavioral stratification could be in chronic disease trials, where diabetes patients may be stratified based on adherence risk using validated questionnaires or digital health monitoring.

And then you have digital phenotyping or AI-driven stratification, which uses advanced

data analytics. Machine learning is used to stratify patients based on complex multi-dimensional data sources in order to uncover hidden patterns in large data sets, automate and scale precision stratification, and enhance recruitment prediction of outcomes or early dropouts. So, some of the data sources that are being used are electronic health records, variable device data, real-world data from insurance claims, mobile apps, and AI-based clustering or prediction models. So, with an example that AI algorithms use electronic health records data to stratify patients with depression based on the risk of relapse, enabling tailored treatment plans for them. Okay, so that was, you know, about the patient stratification.

So the next one is retention. So, it is not only the enrollment of the patients in the trial or the volunteers in the trial, but retaining them is also a big challenge, and that is key to the trial's success as well. So, however, recruitment and retention are expensive affairs. And according to a study, the rates for the same in 2014 to 2015, although it's pretty old data. But it was around, you know, you can see that the patient recruitment is in phase. One was costing around 37,000 US dollars, ranging up to around 300,000 US dollars for the phase four study and 300,000 dollars for the phase three study as well, while retention is also costing a lot, actually.

So, some of the key strategies for patient retention are the challenges we have seen, as well as what those challenges are with retaining the patient. So, now if we look at the strategies to retain patients. So, the first thing is that we need to build trust through communication. So, if we keep patients informed with clear, honest communication to foster trust and long-term engagement. And we can reduce the participant burden as well by offering flexible scheduling and minimizing visits.

To compensate for time and travel and make participation easier, we can use engagement tools like digital apps and regular updates to keep participants connected and motivated. We can manage the adverse events promptly by providing timely medical support and follow-ups to address side effects and concerns early. And we can leverage support networks like we involve patient advocacy groups to offer emotional and peer support, especially in long-term trials. Okay, now where can the AI actually come in? So, as you have seen, AI is transforming not only drug discovery but also drug development. So, AI is transforming clinical trials as well by streamlining patient recruitment, stratification, and retention, and by leveraging ML and NLP, especially natural language processing.

AI quickly sifts through large data sets to identify and match eligible candidates. While also personalizing communication to keep patients engaged. So, the ultimate result of this is faster, more efficient trials with improved data quality and better overall outcomes. So let us see how we can use AI in patient recruitment.

So we can use AI for the screening of patients. So, NLP and ML analyze electronic health records to identify eligible participants faster and more accurately than manual methods. We can do targeted outreach as well as those AI-driven tools that use social media analytics. And online registries to reach specific patient demographics, and we can, you know, by analyzing the large data sets, identify underrepresented populations to improve trial diversity as well. So, we can improve trial diversity as well by using AI in patient recruitment, and there are multiple examples of, you know, those tools people have been using for this task. So, one of the real examples is the IBM Watson and Mayo Clinic.

So they have collaborated to streamline patient recruitment for their clinical trials using AI and ML. Watson Health was able to analyze EHR and other patients' data to identify eligible candidates quickly. And this collaboration has significantly reduced the time required to recruit participants. And since he was able to process vast amounts of data accurately and quickly, researchers were able to focus more on patient care and trial site management than on administrative tasks.

So another initiative was Pfizer's Blue Sky. By integrating AI tools into their recruitment process, Pfizer was able to improve patient identification engagement. Through an AI-driven approach, Pfizer was able to communicate with candidates more efficiently, leading to successful studies. And this initiative has demonstrated how AI can be integrated into large-scale pharmaceutical operations as well. An antidote match, which is, you know, an AI-driven clinical trial matching. So, it's personalized matching using electronic health records, genomics, and lifestyle data.

So it has a faster recruitment process through accurate AI-driven screening and a wide reach with over 250 partner sites and a diverse patient network. And it supports rare diseases by connecting hard-to-reach patients as well. And it's customizable for sponsors, CROs, and healthcare providers. So, talking about the role of AI in patient stratification, using predictive analytics like ML models, AI can predict responses to treatment based on genetic, environmental, and lifestyle data, and this helps refine inclusion and exclusion criteria. It can also help us do the biomarker analysis, where AI can identify biomarkers for precision medicine, enabling trials to focus on patients most likely to benefit from the intervention.

This means we can clearly identify which person will be a responder or a non-responder in this trial. So, a real-world example in patient stratification was adaptive trials by Novartis. So, Novartis used this adaptive trial to demonstrate its commitment to innovation in drug development using AI. And by leveraging adaptive designs, Novartis accelerates timelines, enhances patient stratification, and improves resource allocation while

maintaining high scientific standards. And these efforts contribute significantly to advancing personalized medicine and addressing unmet medical needs efficiently.

Another example is the Mediboost, which is an ML-based tool that builds highly interpretable decision trees with accuracy comparable to ensemble models and is particularly useful for patient stratification in clinical trials where both accuracy and interpretability are essential. Where traditional decision trees are favored for their clarity but often lack predictive power, Mediboost overcomes this limitation by generating single trees that are both accurate and easy to interpret. It has outperformed standard decision trees in 11 out of 13 medical datasets as well. So, this makes it ideal for accurate patient grouping, reduce trial bias, identify subgroups and risk levels and supporting personalized treatment strategies in patient retention.

So we can use it for, you know, predictive retention model. It can identify participants at risk of dropping out by analyzing engagement patterns and medical history. This allows for preemptive interventions like personalized reminders, support, and remote monitoring, especially nowadays with the help of wearable devices. Mobile apps integrated with AI track patient adherence in real time, reducing the burden of in-person visits. And through enhanced engagement using the chat boards or virtual assistants, they provide ongoing support, answer queries, and keep participants informed about the trial's progress. So, these are the ways in which we can use AI in patient retention as well.

An example of using AI in patient retention is Deep6AI. It enhances patient retention by leveraging real-time data and predictive analytics to keep participants engaged in the clinical trial. And by continuously analyzing both structured and unstructured data from electronic health records, the platform detects early signs of potential dropouts and flags them for timely interventions. So, this proactive approach enables the research team to implement personalized communication strategies and adjust trial protocols as needed, ensuring that patients feel supported and valued, which ultimately leads to reduced attrition and more reliable trial outcomes. So, this virtual trial by Medable is another example, which refers to a company's decentralized clinical trial platform that uses digital tools and AI technologies to conduct trials remotely, reducing the need for in-person visits to clinical sites. So Medable's platform is designed to improve patient access, accelerate trial timelines, and enhance data quality by leveraging e-consent, remote monitoring, telehealth, and other digital innovations, leading to increased patient retention.

However, there are some challenges associated with adopting AI in clinical trials, and these are related to, for example, data quality. And standardization, where the clinical data is often incomplete, unstructured, or inconsistent, limits model performance and integration; bias and fairness in AI may reinforce data-driven biases affecting equity. In

patient selection and outcome across populations, interpretability. And trust, because of the black box nature of those models, they lack, you know, transparency, making clinical validation and regulatory acceptance difficult. However, there are, you know, increased efforts to develop explainable AI.

And that may solve this problem in the future, but then you have the regulatory complexity, like the lack of standardized frameworks for AI evaluation. poses hurdle for approval and compliance, workflow integration, embedding AI into existing clinical and trial systems remains technically and operationally challenging. Privacy and ethics raise concerns around informed consent, data sharing, and patient privacy under GDPR, HIPAA, etc. those rules which protect your personal data and generalizability because modern models trained on narrow data sets may not perform well on the diverse population in a real-world setting, the infrastructure and resources, of course, are essential because the implementation of AI requires significant computational resources, secure data handling, and skilled expertise. So, coming to the summary, AI is reshaping clinical trials by optimizing patient recruitment, stratification, and retention through ML, NLP, and predictive analytics.

The AI-powered recruitment uses EHR, social media, and real-world data to identify eligible participants, enhance outreach, and improve trial diversity. Stratification models leverage clinical, demographic, genetic, biomarker, and behavioral data to create precise patient subgroups, improving trial accuracy and therapeutic targeting. The real-world case studies from IBM Watson, Pfizer, Novartis, and Medable highlight the tangible impact of AI on accelerating timelines, improving engagement, and enhancing trial outcomes. However, some of the key challenges, including data quality, fairness, modal explainability, regulatory compliance, and integration, remain a hurdle to the existing clinical workflows.

So in the end, I have an activity for you. Imagine you are part of a startup developing an AI powered platform to support a clinical trial for a new cancer therapy. You must integrate AI for patient recruitment, stratification, and retention. So, what data sources would you use, and how would your AI model make decisions for each stage: recruitment, stratification, and retention? So, I have provided a list of some nice papers that you can go through to get more information about this topic. And with that, thank you.