AI in Drug Discovery and Development

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

Week-08

Lecture-38

Welcome to the course "AI in Drug Discovery and Development." In today's session, we will talk about clinical trial protocol design and optimization, so by the end of this lecture, you will be able to identify the essential elements required to define. Draft an effective study protocol, analyze different clinical study designs, and select appropriate types based on research objectives. Understand the principles behind various blinding methods and their role in minimizing bias, and explore the integration of AI and ML in clinical trial design. and how data-driven tools enhance decision-making, as well as examine real-world case studies demonstrating the impact of AI on protocol optimization. So, let us dive into, you know, the area that is named the clinical trial study protocol.

So, a clinical trial protocol is the essential blueprint for designing and conducting a trial. It outlines key elements, such as the objectives. So, what are the objectives of conducting this clinical trial? What methods are being used? And the endpoints, in the previous session, we discussed the endpoints. So, what those endpoints will be and ensuring all procedures meet regulatory standards like those of the FDA or EMA, or in the case of India, the regulatory agency.

Protecting patient safety and data quality is essential. So, these are created through collaboration among researchers clinicians statistician and regulatory experts and it ensure a comprehensive and scientifically sound design. So, the protocol serves as a standardized guide ensuring consistent execution across all trial sites from recruitment to treatment and data collection. The characteristics of an ideal clinical trial protocol can be defined by this mnemonic called FINER, which is the criteria to maximize the chances for success of a clinical trial. So, these are feasible: the trial should be feasible; it shall be interesting as well, it shall be novel, it shall be ethical, and most importantly, it shall be relevant as well.

So if we look at the, you know, principles of designing a trial protocol. And doing the benefit-risk assessment, you can see this upload where you have the minimum risk and the highest benefits; so, this bottom right quadrant is ours. you know ideal area where the clinical trial design shall lie. So, this is the ideal scenario for phase 2, phase 3, phase 4, or pivotal trials. And then, if you look at this quadrant where the risk is a little bit higher, okay, however, there are benefits, so these are pilot phase one.

or small mechanistic phase two trials; these trials lie in this quadrant where the risk is a little bit higher. Because in phase one we are evaluating the toxicity as well, we do not know whether the drug will show any toxicity. Or not, in that case. So, these are, you know, a little bit on the risky side. And then, on the other hand, where there is no benefit and a lot of risk.

So, in that case, we need to revise the research questions, and we need to try to reduce the risk. And increase the benefit and try to move, you know, whatever is lying here to this corner, which is the ideal situation. And if there is, you know, no risk but there is no benefit as well, so if feasible, cost and practicability-wise, we can run a small trial to optimize the treatment and then try to move. this one into this quadrant. So, our ultimate objective is to bring everything to this quadrant where there is minimal risk and maximal benefit.

So if you look at the life cycle of the clinical trial protocol, it typically spends about 10 to 15 years and can cost anywhere between Half a million to billions of US dollars, depending on the drug's complexity, trial design, and regulatory requirements. So, it starts with the research hypothesis, followed by protocol development, site or trial preparation, subject enrollment, and site selection as well. Data collection, monitoring, processing, and moving to the next step is the trial management and data analytics reporting of the results. And then regulatory submission and finally the drug approval based on the outcome of this clinical trial, so you can see that this is a kind of very long journey. However, based on some specific clauses, it can be reduced for some specific diseases or indications to a shorter period of time, but usually it is quite a longer and really expensive task.

Okay, so let's come to the component of the clinical trial study protocol. So, a protocol synopsis or summary should contain, at a minimum, information on the objectives. So, what are the objectives of the trial? What are the endpoints? What kind of design are you using? And then the schedule of events and assessments, inclusion, exclusion or withdrawal criteria, screening and enrollment, safety assessment, efficacy evaluations, and pharmacokinetic sampling, if involved. So, you know, there are eight steps for drafting a clinical trial protocol. So, the first thing is that you have to understand the requirements or the expectations.

And then we need to set up a protocol development team. And after, you know, setting up a protocol development team, so we need to assign the tasks or deliverables to the team members. We need to draft the protocol synopsis, present the protocol synopsis for review and approval. Draft the entire protocol based on the protocol synopsis, present the entire protocol for review and approval, and release the approved protocol. Okay, so study design, which is, you know, a crucial aspect of clinical trial design.

So the study design is the overall plan for how a clinical study is done, and it explains how participants are chosen, how treatments are given, how data is collected, and how results are analyzed in the trial. So, a good study design helps answer the research questions in a clear and reliable way, improves the quality and trustworthiness of results, and also helps meet ethical and regulatory standards while keeping participants safe. And choosing the right design depends on goals, past research, available data, budget, as well as time. So, if we look at the schema for types of study design, there can be experimental or interventional studies. which can be classified into randomized controlled trials, non-randomized controlled trials, or adaptive trials.

And then there can be known experimental or observational studies if they are descriptive in nature. So, then these are named case series or case reports. Population and ecological studies, if they are analytical in nature, mean we are analyzing the data. and making some, you know, conclusions on the basis of that, so those are usually called cohort studies. A case-control study or cross-sectional study, and another type can be evidence summaries, which are kinds of meta-analyses and systematic reviews.

So if we talk about the quality of evidence in clinical trial design, as we go from bottom to top in this pyramid, the quality of evidence is increasing. And then you can see here, for example, that you have the observational studies, including cohort studies, case series, or individual case reports. And then you have the experimental studies, like randomized controlled trials and non-randomized controlled trials; and then you have the critical appraisals, like meta-analyses, systematic reviews, critically appraised literature, and evidence-based practice guidelines. So, you can see that the quality of evidence is actually going up in this critical appraisal. So, these are mainly based on the evidence.

So, they have, it is like reviewing the previously reported literature and, on the basis of that, making some, you know, conclusions, right? So, the quality of the evidence is increasing if you are going in the upward direction of this pyramid. Okay, the next thing is when we are designing a trial, there is a term called blinding or masking, and that is a very important aspect. This is done to minimize bias in clinical trials by keeping certain individuals unaware of treatment assignments. For example, if a person who is taking the treatment knows that they are getting a medicine or a placebo, it means there is no medicine in that; it is just the excipients. They will have this psychological state of mind that not getting medicine means I will not actually get cured.

So that can affect the outcomes. To minimize the bias in the trial, we actually use this masking, which can be of different kinds. Okay, so it can be an open-label trial where no blinding is done, so all parties, participants, investigators, and assessors know the treatment being administered. It is typically used when blinding is not feasible, for example, in surgical procedures or behavioral therapies; in that case, you cannot use blinding, so it has a higher risk of performance and detection bias. And then you have the single-blind trial where only the participants are unaware of the treatment allocation.

So it helps reduce placebo effects, but observer bias may still be present in this case where the patient or the participant who is taking the treatment does not know whether they are taking the treatment or the placebo. Those are blind to the treatment, actually, and then you have the double blind, where both the participant and the investigator, including those administering treatments and collecting data, are unaware. So, those who are unaware of the treatment allocations are considered the standard for reducing bias, especially in drug efficacy studies. So, this is one of the most popular clinical trial designs, you know, a clinical trial masking, you know, where you use a double-blind trial. Where both the investigators, like those hospital doctors and nurses, are blind to the treatment and the participant is also blind; both of them do not know what.

What is what, who is getting which treatment, actually? And then you have the triple-blind trial, where, in addition to the participants, Investigators, including data analysts or statisticians, are also blinded, which reduces the risk of analytical bias in reporting or concluding the trial results. And then you have a quadruple-blind trial. Blinding is extended to participants, care providers, investigators, and outcome assessors, including the sponsors, which maximizes bias reduction across all stages of the trial and is often used in highly rigorous and multicenter studies. Okay, and then coming to the types of clinical trial designs. So, you know, these designs vary in structure and methodology depending on the research question, intervention type, and feasibility.

So one of the types is randomized controlled trials. So, these are considered the gold standard. you know in clinical research for assessing efficacy and safety of interventions. So, the participants are randomly assigned to two or more groups, typically an experimental group receiving the intervention and a control group receiving a placebo or a standard treatment. So, then this randomization reduces selection bias and ensures comparability between the groups.

It is often conducted as double-blind studies to reduce performance and assessment bias, and it is capable of providing high-quality evidence for causality by minimizing confounding. So, you can see here that the population is split into two groups: one is the control group and the other is the intervention group, along with the outcomes of both groups. These are measured, and here we usually use the double-blind design; they do not randomize or use observational designs. So, these studies do not use random allocation, and participants are actually observed in natural settings. So, these can be cohort studies where they follow groups or cohorts over time based on exposure to a factor, for example, drug or no drug.

These studies can be prospective or retrospective, meaning they can involve analyzing data that has been collected in the past or starting from today, analyzing data in the future. And the case-control studies which begin with the outcome example disease or no disease and look back in time to examine exposure status in that case. And then cross-sectional studies that assess both exposure and outcomes at a single point in time. So, while these studies cannot establish causality definitively, they are valuable for hypothesis generation, long-term outcomes, or when RCTs are unethical or impractical to conduct. So, then you have the parallel group design, where participants are randomized to distinct arms, and each group receives only one treatment throughout the study, and comparisons are made between the groups.

It is simply and widely used, especially when the treatment has a lasting or irreversible effect, and it requires a larger sample size to account for inter-individual variability. So, you do the screening and randomization, give them drug A, drug B, or drug C, and then analyze the results. And then you have a crossover design in which each participant receives multiple interventions sequentially, separated by washout periods to eliminate carryover effects. The participants act as their own control, reducing between-subject variability and increasing statistical power, which is ideal for chronic stable conditions. where the interventions have reversible effects that are not suitable for treatments with permanent effects or where the disease condition may progress rapidly.

For example, screening and enrolling the participants involves having two groups where drug A and drug B are given, and then there is a crossover, meaning that the group that was receiving treatment B will now receive treatment A; however, before that, there is a washout period. So, that you know the remaining drug in this body shall be washed out. So, it means that the same person is being evaluated for different treatments, and then you analyze the results. So, this is called the crossover design.

And then you have a factorial design. So, it tests the independent and combined effects of two or more interventions simultaneously in a single study. So, the participants are randomly assigned to one of the several combinations. For example, you have treatment A and treatment B: yes, no; yes, no. So here in this cell, you will get both treatment A and treatment B; here only treatment A, here only treatment B. And here, there is no treatment that is a control group; okay, there is efficient use of resources as it allows testing for both main effects.

And the interaction effects as well; however, it requires careful planning and a larger sample size if interactions. Effects are expected. So, if there is, you know, interaction between A and B and they are affecting each other's effects. So, then we need to have a, you know, larger sample size. And then you have the cluster randomized trials.

So instead of individual participation, entire groups of clusters, for example, clinics, schools, and communities, are randomized. So, it is useful when interventions are implemented at the group level or when individual randomization is not feasible, and the analysis must account for intra-cluster correlation to avoid inflated type 1 errors. However, it is often used in public health, behavioral, and educational interventional studies. Okay, so after the design, what kind of study are we going to run? Then you can do the master protocol trial designs. So, a master protocol clinical trial includes multiple subgroups and sub-studies with patients having the same or different diseases, and it employs one or multiple interventions or drugs to treat them.

So, it can be classified into, you know, three innovative trial designs: basket trial, umbrella trial, and the platform trial. So, in the basket trial, we test one treatment or treatment combinations across multiple diseases or subtypes that share a common biomarker or genetic features. In Umbrella, we evaluate multiple treatments within a single disease. So, it is kind of the opposite of the basket trial and often involves sub-studies targeting different subtypes of patient subgroups. And the platform trial, where a platform trial assesses multiple treatments for a single disease with the ability to add or remove treatments and sub-studies over time.

Okay, so looking at the importance of clinical trial protocol optimization, optimizing a clinical trial protocol leads to strategic planning, standardized guidelines, and collaborative engagement. Early feedback, as well as reducing the cost, improving efficiency, assessing complexity, benchmarking the trial, and enhancing data quality and regulatory compliance. So, this is an example of the use of optimization in patient recruitment. You use the EHR integration of electronic health records along with the clinical data, which triggers the clinical events, and then you use all this data, for example, NGS results. Provider request and then disease state change or treatment change.

For the new diagnosis, and based on that, you can match the patient to the trial, so patient-trial matching can be from the clinical trial data model. Where you are using all those biomarkers, like protein biomarkers, cytogenetic biomarkers, eligibility checklists, treatment context, or prior therapies, this can also be done on the basis of NGS biomarkers. Once it is done, it will lead to the initial study, initial pre-processing, final pre-processing, consent, screening, and the study management itself. Okay, so now let us talk about what AI brings to the table. AI empowers clinical teams to design faster and smarter.

And more adaptive trials improve outcomes for both patients and researchers, so AI supports key decision-making by providing data-driven insights on trial design and patient recruitment. Site selection helps to expedite the timeline and improve efficiency. While AI is not a standalone decision-maker, it enhances human judgment by offering smart, real-time recommendations. And this synergy creates a paradigm shift in clinical development. And some of the roles that it plays are that it can transform vast data into actionable insights, enhance protocol efficiency and predictive accuracy, and support data-driven decisions from design to execution.

So looking at the AI in protocol design, so it can help us with the literature mining and regulatory guidelines. So, it can mine literature and clinical trials database to identify best practices and integrate current regulatory guidelines. This is done especially by using NLP and predictive analytics for endpoint and eligibility. So predictive analytics can help determine optimal endpoints and eligibility criteria by analyzing historical and real-world data. So, we can use the simulation tools for scenario forecasting, as well.

Like those AI-driven simulation tools, they enable virtual modeling of various trial scenarios to forecast outcomes, refine study parameters, and then generate standardized templates via the NLP. So, NLP can assist in drafting standardized protocol templates, ensuring consistency and clarity across the documents. An integrative ML approach for robust design, where the ML algorithms integrate inputs from clinical, statistical, and regulatory domains to propose robust and scientifically sound trial designs. So, some of the real-world examples are like the Medidata's AI Simulants. So Medidata's generative AI tool creates synthetic data from historical clinical trial datasets, preserving statistical patterns while safeguarding privacy.

And this enabled researchers to simulate trial outcomes and refine inclusion exclusion criteria before finalizing protocol. And the impact of what they have used was that this protocol, you know, drafting time was reduced by almost 50% and minimize post initiation amendments by analyzing close sponsored data from almost 30,000 plus trials. So, you can see that, because of those NLP models, they have a huge capacity to sift through large amounts of data. And then, by using the data, they can gain insight from it and help us design a better clinical trial protocol. And then there is another tool called Clinion eProtocol, where the Clinion AI tool generates a complete clinical trial protocol from the synopsis using a library of historical protocols.

So it automates sections like objectives, endpoints, and statistical plans using TransCelerate templates. And the impact was that it could cut the development time from 160 to 220 hours to just 24 hours, saving at least 2 million annually in operational costs. And then you have the sightline protocol's smart design. So, this AI tool analyzes over 750,000 trials to recommend optimal primary endpoints and inclusion-exclusion criteria.

So, it forecasts the enrollment rate and trial duration based on historical data.

So the overall impact is a 40% reduction in protocol amendments in oncology trials by aligning endpoints with real-world evidence. So, these are some of the, you know, applications of AI in clinical trial protocol optimization. where we can use the AI powered complexity scoring where the scoring models can assess protocol complexity pinpointing areas for simplification to reduce cost and improve efficiency. It can be used for real-time data monitoring and adaptive modifications, facilitating real-time monitoring of trial data and allowing for adaptive modifications to the protocol to enhance operational agility. Continuous regulatory compliance monitoring, where the AI systems continuously compare protocol details against evolving regulatory standards to minimize amendments and ensure ongoing compliance.

and predicting the risk forecasting and mitigation where predictive models can forecast potential risks and such as recruitment challenges or delays that enable proactive mitigation strategies Resource allocation budget optimization, where advanced analytical analytics optimize resource allocation and budget planning by accurately forecasting workload and fiscal requirements. So, this is one of the real-world examples of the use of AI in protocol optimization, where IBM and Pfizer's AI-driven site selection and real-time monitoring work was done. So, they both collaborated to develop an AI model to predict the site performance and enrollment rates. The system analyzes historical site data, patient demographics, and trial complexity to design high-performing sites and flag underperforming ones in real time. So, the impact was to reduce the recruitment timeline by 25% in Pfizer's oncology trials.

and enable multi-dynamic adjustments to site resources cutting costs by almost 15 to 25 million annually and the AI driven next best action recommendations improved retention by 20% through proactive engagement strategies. So, this is another tool: ConcertAI's CTO 2.0. So ConcertAI is a SaaS platform that uses AI to optimize trial design by analyzing almost 900 oncology research sites and over 70 million patient records. So, it integrates EMRs, biomarkers, and social determinants of health to refine inclusion and exclusion criteria and improve diversity.

So the impact is a reduced patient burden of 40% through protocol adjustments aligned with community care standards and enhanced cohort diversity of 35% using AI-driven social determinants of health insights. And it accelerated trial approval by 20% to optimize biomarker-based stratification. And then we have the Lindus Health Citrus platform. So, it's an AI-powered e-clinical platform that automates data monitoring and biostatistics for decentralized trials. So, it uses generative AI to simulate study design and predict enrollment feasibility across 40 million plus electronic health records.

And the overall impact is to reduce the trial timeline by 30% through instant biostatistical analysis. Cut manual data review time by 50% using AI-driven anomaly detection and achieve 95% data completeness in a maternal health study with AI-guided patient engagement. So, looking at the clinical trial design optimization, where it can leverage the quantum computing methods. So, you can see that in the model construction, the study protocol, formulation properties, biological system properties, drug properties, and drug biological properties can be used. Further pre-processing using feature selection, reduction topological data analysis, quantum encoding with the quantum feature map, and model estimation is followed by the model that can model the drug mechanism of action.

So, even this can help in you know clinical trial design and optimization. And that is, I believe, the future of the clinical trials, actually. So, coming to the summary, the clinical trial protocols form the backbone of every clinical study, detailing objectives, methodology, and operational structure. and a well-optimized protocol enhances trial efficiency, improves participant recruitment and retention, and ensures regulatory compliance Selection of the appropriate study design, such as parallel, crossover, adaptive, etc.

It is critical to meet scientific and ethical goals. And blinding and randomization techniques are essential for minimizing bias and improving data integrity. The AI and ML models are revolutionizing protocol design and shaping the future of clinical research by enabling data-driven decision-making and predictive modeling. So, in the end, I have an open question for you. If AI can help design smarter protocols, predict patient dropouts, and optimize endpoints, do you think we are approaching an era in which fully autonomous trials could exist? I have some suggestions for you to learn more about this topic. You can actually go through these papers. And with that, thank you.