**Emerging Programs AI Workgroup Meeting #2**

**Date:** 10/01/2024

**Topic:** Site Selection and Participant Adherence and Retention

## Overview and Current Issues

In the field of clinical research, the growing use of Artificial Intelligence (AI) promises to revolutionize processes like site selection and participant retention and adherence.

Major issues in clinical trials include inefficiencies in site selection, delays in patient recruitment, and difficulties in maintaining patient retention and adherence throughout a trial. Traditional methods of selecting clinical trial sites can be labor-intensive, often relying on expert intuition or small, structured datasets. Participant adherence is a persistent issue, with patients often discontinuing participation or a treatment due to adverse effects, switching medications, or logistical problems (e.g., transportation issues). Clinical research often ignores the social characteristics of patients and populations when assessing the performance of medications and devices in the real world.

Currently, 80% of valuable data needed for decision-making in clinical trials resides in unstructured formats such as clinical records, social media, and patient interactions. AI technologies, particularly Natural Language Processing (NLP), are increasingly being deployed to mine these vast language datasets to extract insights that can optimize clinical trial processes. However, despite its promise, AI's broader adoption faces significant hurdles in terms of economics, culture, regulation, and data privacy concerns. Understanding the current state of AI's role in clinical trials, its challenges, and the steps forward is crucial for researchers looking to leverage these tools effectively.

**Meeting Key Points**:

- AI, and specifically natural language processing (NLP), has potential to streamline clinical trial operations by improving efficiency and accuracy of processes and reducing costs

- AI can analyze social media and electronic medical records (EMRs) to understand patient behaviors and factors influencing retention and adherence
- Whether it's large language models (LLMs), or machine learning (ML), or ontologies, these AI tools solve specific challenges; Sometimes more than one tool is needed to answer a particular challenge
- NLPs have evolved; LLMs and Gen AI continue to expand NLP capabilities
- Despite NLP's potential, the translation from potential to an implemented tool has challenges (i.e. from a research idea to a marketable product that actually works and is cost effective)
- To overcome these challenges, it is critical to:
    - validate the AI models
    - address regulatory and ethical concerns
    - collaborate between AI developers and trial sponsors to ensure that AI solutions are affordable, easily implementable, and tailored to the specific needs of clinical trials

**AI Application Examples for Site Selection and Participant Retention and Adherence**:

NLP can extract valuable information from unstructured text to aid in site selection, participant retention and adherence by helping us find and understand patients. Examples include:

- Site selection: AI can systematically surface information from large databases like clinicaltrials.gov to identify clinical trial sites with the right capabilities
- Participant retention and adherence:
  Sponsor teams typically want to know about switching around adherence, safety, and efficacy
  e.g. Brand product teams want to monitor what patients are talking about around switching their meds. Why? What are the reasons?
    **a.** Is it a dose thing? A regime thing? An adverse event thing? Is it that they've just been talking to friends and other patient groups, and they've heard that a different medicine is more effective?

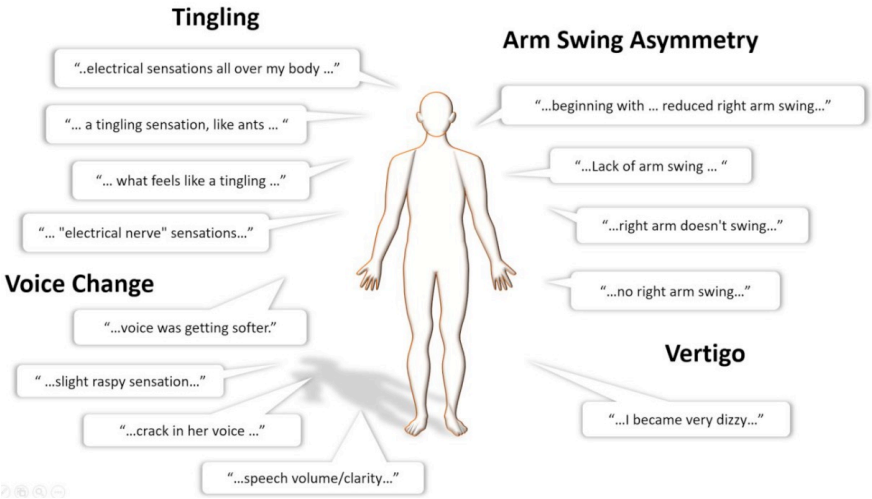
**Case Examples presented:**

1. Site selection: (Merck article)

- Goal: Merck Experimental Medicine division (EMS) needed to locate a clinical trial site for gastric bypass trials with the ability to measure gut peptides before and after surgery.
- Approach: Merck used NLP over their inhouse version of Clinicaltrials.gov and had queries extract terms around gastric bypass, biomarkers, gut peptides, diabetes and obesity, and other key concepts into a structured summary table, which enabled systematic and rapid review of the possible sites
- To assess scientific rank for the investigators, the author and/or the trial ID were extracted from MEDLINE, and results combined
- Search yielded three ideal trial sites; one previously unknown to Merck
- Result: **reduced time burden for clinical trial site planning,** hence **decreasing overall cost**

2. Trial design, patient selection & retention: (Roche article)

- Goal: The preclinical and early development teams within Roche wanted to understand how patients were describing their disease to support patient-centered drug development.
- Approach: They used a combination of AI tools for social media to get a bottom up view of what patients think are important to inform the right endpoints for early Parkinson's disease.
- used NLP on “highly trusted” social media sources (i.e. patient-focused communities rather than more diverse Twitter or Facebook posts) to categorize comments into symptom or impact categories
- Result: NLP extracted symptoms and impacts confirmatory of the clinical trial endpoints
    - In addition, discovered new signals (e.g. tingling, voice change, vertigo)
    - These recommendations were taken forward with clinical teams for additional patient-centric clinical endpoints




3. Retention and Adherence: (North Shore and IQVIA example)

- Goal: NorthShore University Health System combined with Edward-Elmhurst Health wanted to gain a greater understanding of the health equity disparities that exist within their populations, and understand whether some of the reasons why some patients might leave a clinical trial or stop taking their medicines are because of social and economic issues (e.g. transportation issues fall into social determinants of health)
- Understanding social determinants of health and how, at a population level and at a patient level, would help make sure patients are being cared for in an optimal and equitable way.
- The challenge is identifying the amorphous concepts from the complexity of unstructured data (from a clinical trial perspective, pharma companies don't always have access to the kind of detailed information that a healthcare provider has access to.)
    - Patient level SDOH data was limited to poorly populated structured data within the EMR, while research shows that SDOH factors are documented in over 30% of patients.

- Approach: NorthShore used IQVIA's SDOH NLP in the ED where it screens free text clinician notes and flags patients with SDOH risk factors, enabling more targeted triage of at-risk patients
- Result: ED social workers assessed 56% more patients for SDOH needs that would not have otherwise been identified
- Impact: Translating that to trials, being able to surface the information around transportation issues, ambulatory status, mental health and so forth, is valuable in retaining patients.

**Challenges & Solutions to Adoption:**

| **Challenges** | **Solutions** |
|---|---|
| Bias<br>e.g. reporter bias- there is a specific type of patient that winds up posting to social media | To tackle biases, recognize where those biases come from<br><br>Need to assess where and what type of data might be most reliable to help understand a question of interest<br><br>e.g. social media, because it generally tends to be a very broad landscape compared to people who fill in surveys... it may remove some of the biases that we might otherwise get in clinical trials |
| Expertise<br>e.g. Domain-specific expertise to fine tune AI models for effectiveness in specific tasks | Medical scientists are needed to curate and validate.<br>Test the extraction rules used for each tool<br>Have some level of human in the loop |
| Continuous data validation<br>(e.g. to address concerns over hallucination) | Have standardized validation protocols<br>Have frameworks that evaluate the trustworthiness and effectiveness of AI models in clinical settings |
| Data privacy and transparency in decision-making processes | Have a clear informed consent to ensure that patients are aware of how AI is being used in their care.<br>Support transparent and fair AI algorithms |
| Uncertainty about regulator acceptance | Be able to describe how and where AI is being used, whether there is any risk to participants, and mitigation of those risks |
| Costs and operational challenges<br><br>e.g. integration of AI with existing systems<br><br>AI tools often require substantial upfront investment, not only in terms of financial resources but also computational power | Developers need to offer something that is a validated service, affordable, implementable and sustainable.<br><br>Buyers (e.g. Sites or sponsors) need an implementation plan including a way to assess |

| | |
|---|---|
| and expertise. Many research institutions may lack the infrastructure or financial capacity to implement and maintain AI-driven solutions | if the tool is affordable and does what is necessary. |

**Perspectives:**

Sponsors note the need for:

1. understanding the differences between rules-based approaches, machine learning, and large language models to select the right tool for the task.
2. frameworks to evaluate the trustworthiness and effectiveness of AI models in clinical settings
3. having processes and technologies in your workflow that can find new information
4. generalist models (like GPT-4) as they can outcompete specially-tuned models like Med-PaLM, Microsoft article https://arxiv.org/abs/2311.16452

Regulators comment on:

1. aligning AI technologies with current regulations
2. being aware of the potential for AI to introduce biases
3. the need to describe: 1) How are you going to use the tool? 2) Where is it going to be used? (i.e. upstream in your drug development? Or closer to the trial design? 3) Is it going to be interacting with the participants of your trial?
4. the importance of ongoing monitoring and evaluation (the closer you get to the patients or the participants with AI, the more you are subjecting them to risk, then the scrutiny to your modeling approach is going to be higher)
5. that FDA is using AI for post-market safety surveillance and looking at adverse events, using social media but there's a 100% verification with human in the loop right now to see if this is a good use of AI

Sites question:

1. Are we replacing human cost and efficiency cost with computational cost to be implementing all these models?
2. Is the evaluation process to validate algorithms and interface with our electronic health records equally/successfully transferable across applications (e.g. understanding Parkinson's disease endpoints vs finding patients for a cancer clinical trial)?
    a. every clinical service that we're rolling out we have to do a validation exercise, and then fine tune, the model for that particular population.
3. How can we get access to a tool that's affordable and does what is necessary for that context?
4. To what extent do you use the tool as is then discard versus building a technology that's fit for a particular purpose?
5. Is there a risk of AI perpetuating existing biases (e.g. leading to concerns for the potential impact of AI on equity in patient recruitment)?

**Future Directions**

1. **Invest in Training and Capacity-Building**: Researchers, clinicians, and trial sponsors need training on AI technologies to understand how to leverage them effectively and ethically.
2. **Conduct Pilot Projects and Validation Studies:** AI tools should be selected based on their suitability for specific trial challenges. Researchers should assess the needs of the trial and test the practical applications of AI in real-world clinical trials to gather insights for further refinement and improvement. AI models should be regularly validated and tested for accuracy, reliability and fairness in different clinical trial contexts.
3. **Mitigate Bias:** Develop methods to identify and mitigate biases in AI algorithms to ensure fair and equitable outcomes. Incorporating human oversight throughout the process helps mitigate risks, such as bias or "hallucinations" (false predictions), and ensures trust in AI-driven results.
4. **Ensure Transparency:** Create transparent and fair AI algorithms. Developers should work closely with trial sponsors and regulators to ensure that AI solutions are practical, effective, and aligned with regulatory requirements. Clear processes must be in place for how AI tools are integrated into clinical workflows. This includes transparent protocols around patient consent, data privacy, and how AI models are used to make decisions.
5. **Focus on Ethical AI**: Address ethical concerns, such as patient equity and bias in data. AI algorithms should be designed to be transparent, fair, and monitored for ethical issues throughout their use in clinical trials. We should also understand in what cases the use of AI didn't work and why.
6. **Develop Standards:** Develop standardized validation protocols and guidelines for AI implementation in clinical trials. Ensure ongoing monitoring and evaluation of AI technologies to address potential biases and ethical concerns.
7. **Monitor Development and Implementation**: Stay aware of new AI models, improvements in computational resources, and changes in regulatory guidelines. As AI tools get more powerful, and you have more data to work with, it’s more likely general tools can do task A and task B and task C. We should also monitor knowledge base and whether we are maintaining knowledge when relying on technology as well as how these tools change the way we do research?