Emerging Programs AI Workgroup Meeting #4

Date: 05/08/2025

Topic: AI Applications in Clinical Data Collection, Management, and Endpoint Assessment

Overview and Current Issues:

The fourth convening of the Emerging Programs AI Workgroup marked the completion of discussions aligned with Section C of the FDA's 2023 AI Discussion Paper. This convening focused on the evolving role of AI in clinical trial data collection, management, and endpoint assessment — areas that are becoming increasingly central to trial execution and oversight.

The discussion included opportunities to increase efficiency in data collection, challenges related to implementation, such as navigating regulatory uncertainty, ensuring responsible use of synthetic data, and shifting entrenched workflows. Participants conversed over the broader implications of data governance, ethical use, and the shifting regulatory and operational landscape in the U.S. and abroad.

Meeting Key Points:

- **Applications for Data Collection and Management:** AI for eCRF automation, audit trail review, image analysis, data reconciliation, and medical coding are in use to improve speed and consistency of operations in clinical trials.
- **Synthetic Data for Testing, Trial Design, and Control Arms:** AI-generated synthetic data is increasingly used to predict outcomes and improve trial design by simulating patient journeys, testing trial systems, and constructing synthetic control arms. As use expands, it's critical to differentiate between synthetic control arms, external controls, and digital twins—each with distinct regulatory and ethical implications.
- **Implementation Barriers:** AI adoption remains limited by organizational inertia, unclear ROI, and misaligned business models. Legal constraints, consent requirements, and governance for data reuse and AI training add complexity. Adoption has progressed faster in clinical care (e.g., ambient voice) than in research, where workflow integration and validation remain challenging.
- **Future Directions:** Progress will depend on clearer standards for synthetic data, stronger consent and governance frameworks, and better organizational readiness. Bridging

clinical and research use, clarifying AI use cases, and enabling precompetitive data sharing will be key to scalable, ethical adoption.

Case Examples presented:

Guest speaker Lisa Moneymaker (Medidata) shared practical examples of how AI is currently being applied to streamline clinical trial operations and support endpoint assessment. Her presentation covered tools in active use as well as forward-looking applications under development or in pilot testing. She discussed:

- **Accelerated eCRF Buildouts:** AI tools are used to ingest protocol documents and auto-generate electronic case report forms (eCRFs), including suggested edit checks and custom functions.
- **GenAI-Powered Audit Trail Review:** A newly released solution enables natural language querying of audit trail data, helping sponsors surface patterns of interest—such as data entries outside expected time windows or unusual behavior at specific sites. This improves both compliance oversight and speed of review.
- **AI in Medical Imaging and Endpoint Assessment:** In collaboration with Dassault Systèmes, Medidata is deploying AI tools to automatically assess tumor displacement from imaging data. These tools are integrated directly into the eCRF and can support multi-organ assessments, enhancing radiologic endpoint review capabilities.
- **Data Reconciliation and Anomaly Detection:** AI systems identify discrepancies across adverse event, concomitant medication, and medical history datasets. These tools use machine learning to flag records for human review, helping trial teams prioritize areas requiring adjudication during data cleaning.
- **Medical Coding with Confidence Scoring:** Medidata's AI-driven automatic coding tool predicts MedDRA and WHO Drug terms and assigns confidence scores to each suggestion. This speeds up the coding process, reduces manual workload for clinical staff, supports consistent, high-quality term matching across large datasets, and improves efficiency during database lock preparation.
- **Synthetic Data Simulation and Control Arms:** Medidata's "Simulants" platform generates synthetic patient data for system testing, model development, and clinical trial simulation. The platform has supported FDA-engaged synthetic control arms in therapeutic areas such as glioblastoma, acute lymphoblastic leukemia, and ovarian cancer.

Challenges to Adoption:

Challenges	Description
Organizational Inertia	Many teams lack change management strategies to support adoption of AI. New tools require users to shift workflows and roles, which can lead to resistance.
Lack of Baseline Metrics	Sponsors often don't track key performance indicators (KPIs) like eCRF build time or coding throughput, making it difficult to measure AI's added value.
Misaligned Incentives	Traditional CRO business models, which rely on manual labor, may disincentivize process automation and efficiency gains from AI integration.
Regulatory Ambiguity	Current frameworks do not clearly define expectations for AI validation, synthetic data reuse, or how to structure informed consent for secondary data use.
Data Governance Complexity	Proprietary trial data is protected by IP and privacy regulations, limiting the availability of real-world data for model training and validation.
Cultural and Educational Gaps	Confusion between different AI applications (e.g., synthetic control arms vs. digital twins) and lack of AI fluency can hinder cross-functional alignment.

Perspectives:

Sponsors and Trial Teams:

- See potential for AI to improve efficiency in coding, data review, and system testing.
- Face challenges demonstrating ROI due to limited baseline metrics and integration hurdles.
- Seek clearer guidance on the use and acceptability of synthetic control arms.

Regulators:

- Emphasize the need to distinguish between synthetic control arms, external controls, and digital twins.
- Cite data-sharing constraints tied to IP protections and trade secrets.
- Note growing interest in policy updates to support responsible secondary data use.

Technology Developers:

- Report strong technical capabilities but slow adoption due to organizational resistance.
- Highlight demand for AI tools without corresponding investment or change readiness.
- Point to synthetic data as a key enabler for testing, training, and trial augmentation.

Academic Experts:

- Observe faster AI adoption in clinical care than in research settings.
- Stress the importance of education and transparency to build trust.
- Call for alignment between ethics, operations, and regulation to support responsible use.

Future Directions:

1. **Change Management Support:** Equip teams with adoption strategies, onboarding plans, and operational guidance to transition from traditional processes to AI-augmented workflows.
2. **Advance Synthetic Data Standards:** Clarify appropriate use cases for synthetic data, including testing, simulation, and control arms, and define validation practices in collaboration with regulators and IRBs.
3. **Improve AI Adoption Readiness:** Encourage organizations to establish baseline KPI tracking and metrics to evaluate the impact of AI tools and demonstrate time or resource savings in measurable terms.
4. **Realign Business Models:** Encourage innovation among CROs and service providers by linking contracts and performance incentives to outcomes, not headcount.
5. **Invest in Cross-Functional Education:** Foster shared understanding across clinical, statistical, regulatory, and operational teams by clearly defining AI terms, roles, and risks.
6. **Clarify Regulatory Expectations.** Explore standardized informed consent language, transparency standards, acceptable use cases for AI-generated and synthetic data, and frameworks to enable ethical, scalable secondary data use and model development.
7. **Promote Data Sharing Frameworks:** Explore public-private partnerships and precompetitive collaborations to unlock access to high-quality, de-identified trial data and to expand model training datasets while preserving IP and patient privacy.
8. **Differentiate AI Use Cases:** Promote clearer definitions and education around synthetic control arms, digital twins, and external controls to support appropriate application and regulatory alignment.
9. **Bridge the Research–Care Divide:** Translate successful AI use cases from clinical care into research contexts, adapting for trial-specific needs and compliance requirements.