Thermo Fisher SCIENTIFIC

CMO 360 Summit

How is the Clinical Trial Industry Leveraging AI in Clinical Development?

Peter G. Linde, MD, FASN – VP, Head Global Medical Officer
Biotech Solutions
Thermo Fisher Scientific/PPD

Acknowledgements
- Rakhi Kilaru
- John Van Hoy

Framing Unsolved Problems in Clinical Trials – How can AI help?

- ***Clinical trials are essential for demonstrating the efficacy and safety of potential treatments***

- ***Over 86%of drug trials and 96% of oncology trials fail to achieve FDA approval***

- ***Contributing factors (High Level):***

  Flawed study design

  Poor patient recruitment and retention/adherence

  Protocol deviations→ Regulatory non-compliance/invalidation

  Insufficient real-world applicability in the intent-to-treat population

Benbow et al. *Harnessing Artificial Intelligence to Transform Clinical Trials and Cancer Care: Opportunities and Challenges* ***Cancer J 2025;31:e0796***)

Framing Unsolved Problems in Clinical Trials – How can AI help?

By the Numbers……

- ***Recruitment delays affecting 80% of studies***
- ***Escalating costs exceeding $200 billion annually in pharmaceutical R&D***
- ***Success rates below 12%,***
- ***Data quality issues affecting 50% of datasets***

Olawade et al. *Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges, and future directions* International ***Journal of Medical Informatics 206 (2026) 106141***

Current AI Framework in Clinical Trials Industry

- *Design & Feasibility (better protocols, realistic cohorts, fewer amendments)*
- *Site & Investigator Strategy (faster selection, better fit, enrollment realism)*
- *Recruit & Enroll (eligibility matching, screening efficiency, diversity/representativeness)*
- *Run & Engage (adherence/retention, decentralized execution, patient support)*
- *Monitor & Assure (RBM, anomaly detection, quality and compliance risk control, adjudication)*
- *Report & Disclose (regulatory documentation e.g. CSRs transparency/disclosure automation)*

Today's SELECTED AI-In-Clinical Trials Use Cases

Retention/ Adherence

Anomaly Detection

Clinical Event Adjudication

Limitation of AI Use in Improving Retention

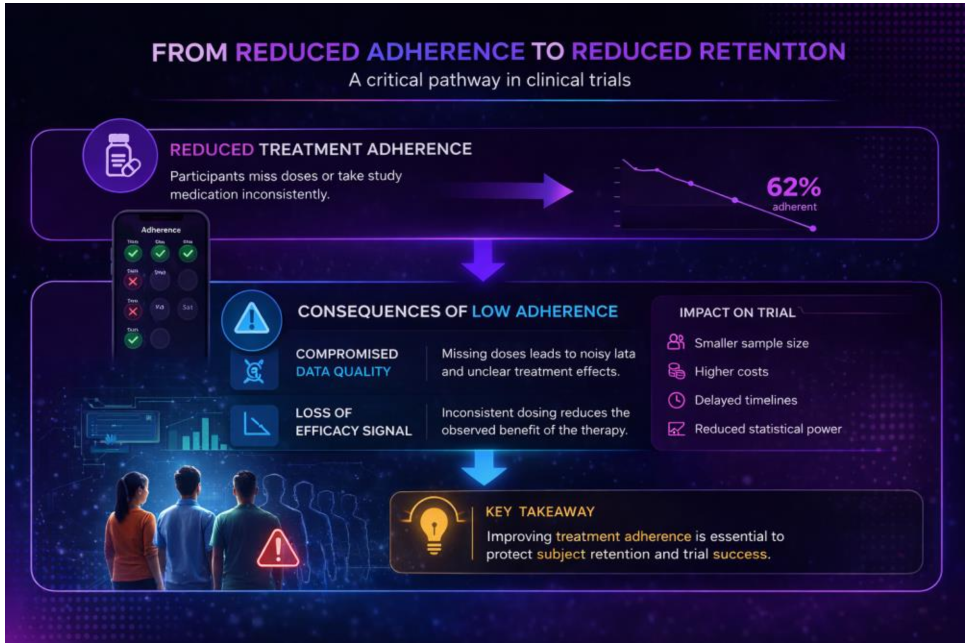
Thermo Fisher SCIENTIFIC

With Respect to AI's Ability to improve Retention, Perhaps the Jury Is Still Out….

- ***Low number of publication of AI-supported Retention Use Cases in the literature***
- ***Many reported studies are based on pilot, simulated or theoretical models rather than large-scale randomized controlled trials, raising questions on generalizability and real-world performance***
- ***Limited availability of required specialized knowledge in AI systems to promote responsible use and broader appeal and adoption***

HOWEVER,

- *A critical "feeder problem" into reduced Retention is **reduced Adherence***
- *AI-supported predictions on Adherence are available ---AI's potential usefulness in this critical trial success measure*



CMO 360 Summit

Selected Use Cases – ADHERENCE

- ***Patients were recruited from two Phase II randomized, double-blind, placebo-controlled trials evaluating the efficacy, safety, and tolerability of orally administered treatment during a 28-week treatment period.***
    - ***In the schizophrenia trial, Patients (N = 264) were randomized to 2 doses of treatment or placebo, administered daily for relapse prevention of schizophrenia until the trial end.***
    - ***In the attenuated psychotic disorder trial,(N = 50), the primary endpoint was time to first episode of psychosis until trial end.***
- *In both trials, treatment adherence was monitored by a computer vision-assisted smartphone application, which records the medication ingestion event, and traditional pill counting methods.*

Zhu et al, AI-based medication adherence prediction in patients with schizophrenia and attenuated psychotic disorders, Schizophrenia Research 275 (2025) 42–51

Selected Use Cases – ADHERENCE in 2 Neuro Trials

Thermo Fisher
SCIENTIFIC

Table 1
Evaluation metrics for the AiCure models using different monitoring periods (Experiment 1), adherence cut-offs (Experiment 2) and trial timepoints (Experiment 3).

Predictive model type	AUC	Accuracy	Precision	Recall	FOR	FNR
A: AiCure models using different monitoring periods						
7 day	0.77	0.70	0.77	0.66	0.37	0.34
10 day	0.79	0.72	0.77	0.70	0.33	0.30
14 day	0.81	0.74	0.81	0.69	0.33	0.31
B: AiCure models using different adherence cut offs						
Cut off 0.6	0.87	0.85	0.90	0.90	0.32	0.10
Cut off 0.7	0.85	0.81	0.90	0.82	0.36	0.18
Cut off 0.8	0.81	0.74	0.81	0.69	0.33	0.31
C: AiCure model at different trial timepoints						
Trial start	0.87	0.85	0.90	0.90	0.32	0.10
Mid-trial	0.85	0.90	0.92	0.95	0.20	0.05
Trial end	0.92	0.87	0.93	0.90	0.30	0.10

Area under the curve (AUC), accuracy, precision, recall, false omission rate (FOR) and false negative rates (FNR) values from Experiment 1 evaluating the 7-, 10-, and 14-day medication-adherence monitoring periods at the 0.8 cut-off threshold for adherence (A); Experiment 2 evaluating 0.6, 0.7, and 0.8 adherence cut-off thresholds for adherence using a 14-day monitoring period (B); and Experiment 3 evaluating the effects of different monitoring timepoints throughout the trial (C) on adherence prediction.

Adherence prediction model developed using the XGBoost tree-based machine-learning algorithm (AI MODEL) in this study can be successfully applied INSIDE THE TRIALS.

RESULTS:

- **14-day period provided more optimal predictive accuracy than the shorter (7- and 10-day) periods during the 28-week trials.**
- **Predictive value is mildly dependent on the adherence cut-off and timing of monitoring during the trial**

Therefore, data had to be analyzed at later timepoints and at a lower Adherence cut-off in order to be slightly more accurate.

Zhu et al, AI-based medication adherence prediction in patients with schizophrenia and attenuated psychotic disorders, Schizophrenia Research 275 (2025) 42–51

Selected Use Cases – ADHERENCE in 2 Neuro Trials

Table 1
Evaluation metrics for the AiCure models using different monitoring periods (Experiment 1), adherence cut-offs (Experiment 2) and trial timepoints (Experiment 3).

Predictive model type	AUC	Accuracy	Precision	Recall	FOR	FNR
A: AiCure models using different monitoring periods						
7 day	0.77	0.70	0.77	0.66	0.37	0.34
10 day	0.79	0.72	0.77	0.70	0.33	0.30
14 day	0.81	0.74	0.81	0.69	0.33	0.31
B: AiCure models using different adherence cut offs						
Cut off 0.6	0.87	0.85	0.90	0.90	0.32	0.10
Cut off 0.7	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]
Cut off 0.8	0.81	0.74	0.81	0.69	0.33	0.31
C: AiCure model at different trial timepoints						
Trial start	0.87	0.85	0.90	0.90	0.32	0.10
Mid-trial	0.85	0.90	0.92	0.95	0.20	0.05
Trial end	0.92	0.87	0.93	0.90	0.30	0.10

Area under the curve (AUC), accuracy, precision, recall, false omission rate (FOR) and false negative rates (FNR) values from Experiment 1 evaluating the 7-, 10-, and 14-day medication-adherence monitoring periods at the 0.8 cut-off threshold for adherence (A); Experiment 2 evaluating 0.6, 0.7, and 0.8 adherence cut-off thresholds for adherence using a 14-day monitoring period (B); and Experiment 3 evaluating the effects of different monitoring timepoints throughout the trial (C) on adherence prediction.

Zhu et al. AI-based medication adherence prediction in patients with schizophrenia and attenuated psychotic disorders, Schizophrenia Research 275: (2025) 42–51.

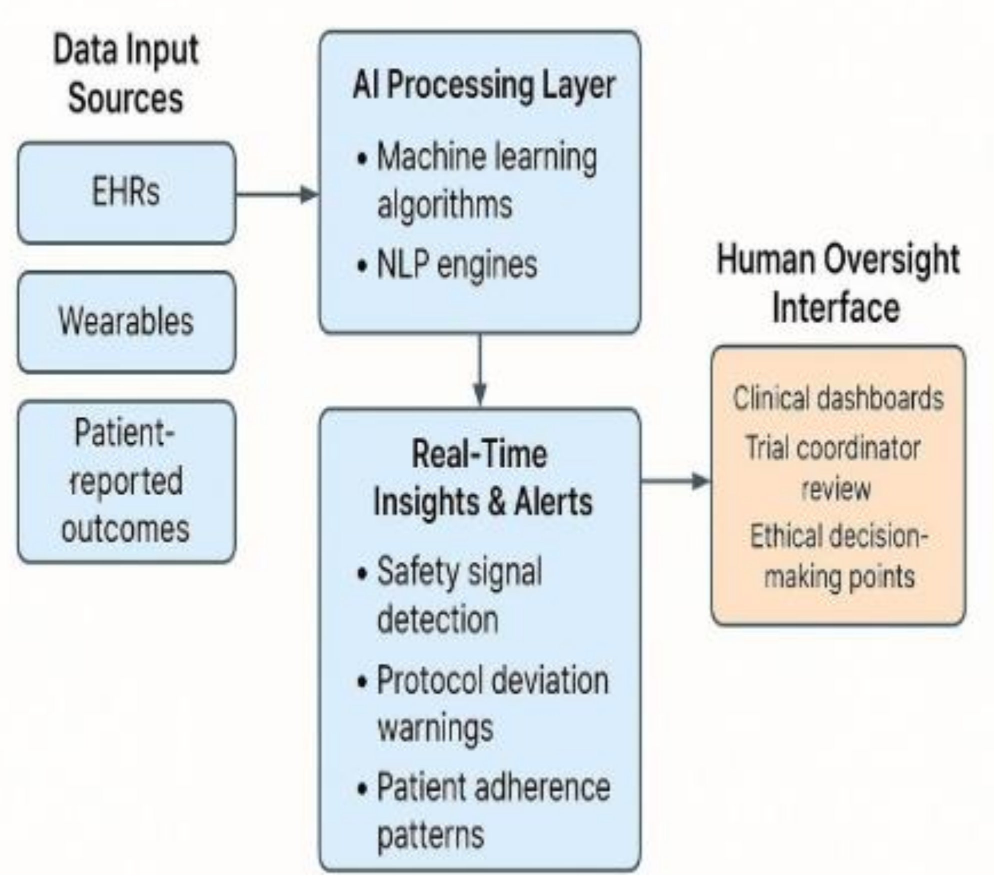
ADDITIONAL RESULTS:

- **A reduction in risk of first relapse was observed with BI 409306 50 mg versus placebo <u>when only adherent completers (≥80 % across trial) were considered</u> (HR [95 % CI] = 0.485 [0.162, 1.452]).**
- **Strong Trend in Reduction in risk was observed in patients <u>who were predicted to be ≥60 % adherent based on the first 2 weeks of trial data (HR [95 % CI] = 0.510 [0.225, 1.156])</u>**

Potential Impact: If In-study Early Predicted Adherence falls below certain thresholds, "in flight" corrective actions can be taken to increase study POS.

Selected Use Cases – Anomaly Detection- Introduction

The Role of AI in Real-Time Data Analytics
for Clinical Trial Monitoring[2]



2- Parasaram, V. K. B. (2025). Real-time clinical trial monitoring with AI-powered analytics. *Journal of Advances in Pharmaceutical Sciences, 3*(2), 36–47.

3- Olawade DB, et al. Artificial intelligence in clinical trials: A comprehensive review of opportunities, challenges and future directions. Int J Med Inform. 2026

Selected Use Cases: Anomaly Detection Example

Case Study:

- ***A completed studies of a novel VEGF inhibitor in 1530 patients with metastatic colorectal cancer. Data were collected at 184 investigator sites, yielding 7208 on-treatment visits and 56 numerical features (data categories).***

  - ***Alkaline-phosphatase, an established hepatic-safety biomarker, was chosen as the key outcome to be monitored.***
  - ***The remaining 55 features (including site and patient descriptors, adverse-event counts, exposure metrics, concomitant-medication counts, other laboratory parameters, and vital signs) were monitored simultaneously as Inputs to capture potential early safety signals, data-quality issues, or other systematic errors.***

Reference: *Yin S, et al. Assumption-Agnostic Deep Learning Framework for Holistic Monitoring. Therapeutic Innovation & Regulatory Science. 2026.*

Selected Use – Anomaly Detection with 2 AI Models

ThermoFisher SCIENTIFIC

Simulated aberrant site prevalence "Infused into the Model": 10%, 20%, 30%

4 Scenarios of Differing Co-Variants used for each % Aberrant Site Prevalence level

LSTM autoencoder framework delivers the best overall and most balanced results

Anomaly	AI Model #1: Deep Learning Prediction Model Long Short-Term Memory Performance	AI Model #2: Classical Machine Learning Performance
Low-prevalence signal detection *10% Aberrant site prevalence*	**F1: 0.46 – 0.77** Sensitivity: 0.5 - 1 Precision: 0.43 – 0.71	**F1: 0.16 – 0.6** Sensitivity: 0.17 - 1 Precision: 0.12 – 0.5
Moderate anomaly burden *20% Aberrant site prevalence*	**F1: 0.71 – 0.88** Sensitivity: 0.83 – 1 Precision: 0.63 – 0.85	**F1:0.31 – 0.89** Sensitivity: 0.83 – 1 Precision: 0.19 – 0.8
Higher anomaly burden *30% Abberant site prevalence*	**F1: 0.81- 0.92** Sensitivity: 0.89 – 0.94 Precision: 0.71 – 0.89	**F1:0.42 – 0.92** Sensitivity: 0.56 - 1 Precision: 0.33 – 0.89

- Performance metrics: F1: Harmonic mean of precision and recall (balances precision and recall); TPR: Sensitivity; PPV: Precision. Higher values better
- Sensitivity focus: Ability to detect true risk, Precision focus: Avoid false alarms. **F1s over 0.7 and higher would be considered "good performance"**

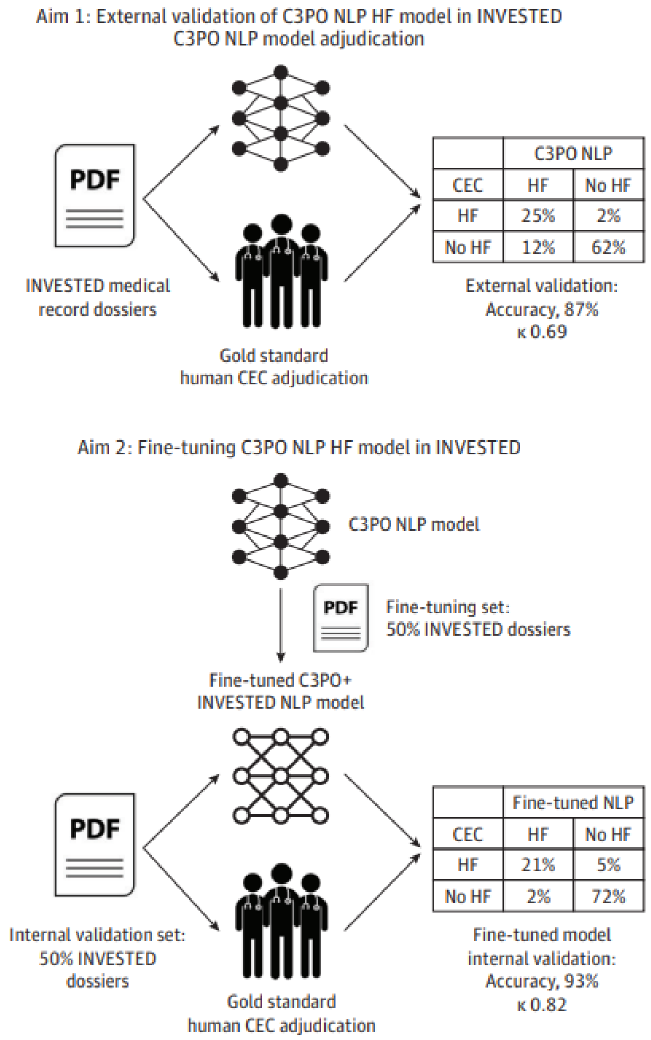
Reference: Yin S, et al. Assumption-Agnostic Deep Learning Framework for Holistic Monitoring. Therapeutic Innovation & Regulatory Science. 2026.

Anomaly Detection: Key Limitations

- **Prediction**: Models that generate ‘expected normal’ patterns use classical predictive learners and deep learning algorithms. In real trials, misconduct, anomalies already exist in the historical dataset, if biased behavior is embedded in training data, the model would learn it as ‘normal’
- **Limited Data availability:** Current limitations around access to large (public) volume of data to train the models. Synthetic datasets a possibility but subject to utility/privacy trade-offs
- **Clinical vs. Statistical Relevance:** For example, clinically severe but rare SAEs may not be detected (omission of SAE or misclassification of severity also can confound the model).
- **High expertise burden:** Architecture choice (of neural network), hyperparameter “coefficient” tuning etc. needs specialized AI-model creation skills
- **High Computational Cost:** Training inference can be slower with non-parallel computing setup
- **Transferability/Generalizability:** Transfer limitations given limited generalizability to the same vs different Indication/therapeutic training dataset contexts

Reference: Yin S, et al. Assumption-Agnostic Deep Learning Framework for Holistic Monitoring. Therapeutic Innovation & Regulatory Science. 2026.

Selected Use Cases – CVOT Trial – AI- Based Adjudication Example using NLP



CEC indicates clinical event committee; C3PO, Community Care Cohort Project; HF, heart failure; NLP, natural language processing.

- **C3PO NLP Model** compared to a gold-standard CEC in a multicenter NIH–sponsored RCT, the Influenza Vaccine to Effectively Stop Cardio Thoracic Events and Decompensated Heart Failure (INVESTED) study.
- **INVESTED Trial** enrolled **5260 patients** with recent acute myocardial infarction or HF hospitalization. Participants received **high-dose trivalent influenza vaccine or standard quadrivalent influenza vaccine**
- The Model showed **substantial agreement** (raw agreement **Initial Accuracy was 87%** [95% CI, 86-88].
- **Fine-tuning the C3PO NLP model** using half the INVESTED hospitalizations significantly improved agreement with the gold-standard CEC adjudication. **Accuracy increased to 93%.**

JW Cunningham et al. JAMA Cardiol. 2024;9(2):174-181

Thermo Fisher Examples

Thermo Fisher SCIENTIFIC

Clinical Trial Forecasting Suite (CTFS)

Solution Approach

The Clinical Trial Forecasting Suite is a proprietary, ***AI-powered platform*** designed to augment FSO delivery partnerships and enhance clinical trial planning precision, integrating real-time clinical, operational, and historical data.

Patient Enrollment + Study Milestone Probability

Site Performance Insights + Early Risk Indicators

Real-Time Collaboration + End-to-end Scenario Sharing

POWERFUL TODAY

18 Therapeutic Areas

219 studies supported since April 2025

143 studies supported since April 2025

Pre-award → **Feasibility** → **Start-Up** → **In-Flight**

Global site and country strategies

Milestone & delivery progression indicators

AI Chat Interface

Site intel & list development

New site & delivery notifiers

Site Activation + Enrollment forecasting & in-flight revisions

Scenario risk early indicators & cross-scenario comparisons

Performance Metrics:

We conducted a retrospective simulation using actual site enrollment data from 156 completed studies

When the AI models are used to replace up to 25% of predicted-low-performing sites, we calculated the following reductions in the enrollment phase:

- **-median across all TAs- 28-day reduction**
- **-median for GI studies: 44-day reduction**
- **-median for Onc studies: 34-day reduction**

Scientific Surveillance (Statistical Modelling - 'AI Convertible')

End to end curation with 'human-in-the-loop, subject matter expertise' and actionable insights

Transformed Data

↓

Unsupervised and semi-supervised anomaly detection
Bayesian Analyses
Predictions
Statistical Process Control
Robust statistical methods
Rule-based systems

↓

Actionable, Explainable Anomaly Reports

Use cases of retrospective and prospective application of scientific surveillance:

Alzheimer's Disease with moderate to severe psychosis phase 3 trial: N=360

- **Results from Blinded Multivariate Analysis & Stat process control detected anomalous improvements** with NPI-C:H+D and clinical anchors and inconsistencies in rater assessments **at one particular site.**
- *Upon integrated review against all other findings, consistent delay in AE reporting was also reported at this same site.*
- **This initiated multiple corrective actions spanning rater retraining and removing a rater consistently erratic in their assessments at this site**.

Atopic Dermatitis phase 2 trial: N=396

- **Results from Blinded Multivariate Analysis & Stat process control detected directionally inconsistent results** on similar constructs (e.g., itch, BSA) from EASI, SCORAD, clinical and patient anchors relative to all other inconsistencies within and between scales (including PROs).
- **This initiated multiple corrective actions spanning rater training, newsletters, tip sheets, reminders were issued resulting in successful interim analysis readout displaying stat significant results**

Integrated Surveillance

Medical, Statistical and Clinical Science Reviews

Protection of Scientific Integrity

Operational Corrective Actions

Key message

In-Flight Blinded Integrated surveillance with clinical, statistical and medical science insights are highly effective to maintain Rater Consistency in ongoing trails

RECAP: Current Efforts to Establish Value of AI in Clinical Trials

Thermo Fisher SCIENTIFIC

- AI's Leverage Points Cluster into **Five Pillars**:
  - **Cycle-time compression** (start-up, screening, monitoring)
  - **Cost-to-deliver reduction** (virtual/decentralized execution, automation)
  - **Data quality and integrity** (accelerated data cleaning, data anomaly detection)
  - **Enrollment performance + representativeness** (enhanced subject matching/recruitment), minimizing missed eligibility)
  - **Risk and compliance management** (Risk Based Monitoring, AE signal identification)