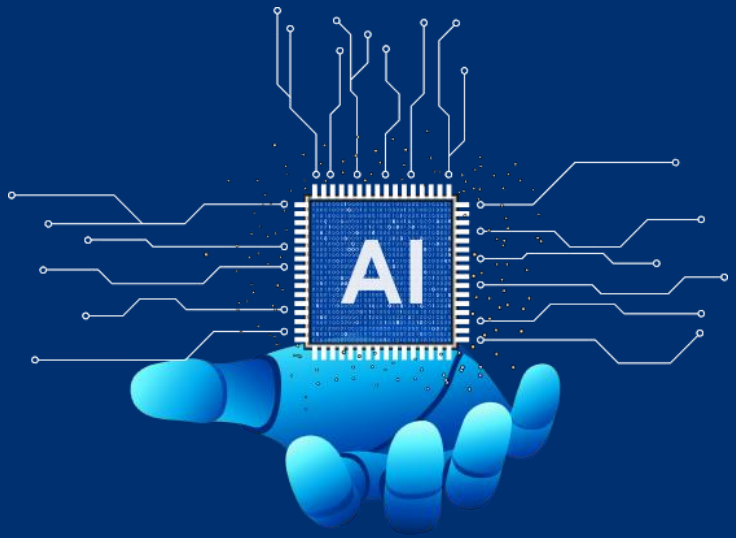

phuse
connect·share·advance
External Control Arms in Clinical Trials
with AI-Enhanced Analytics
PHUSE Mumbai SDE 2026
phuse.global

Agenda



1. Emerging Trends in Data-Driven Clinical Trial
2. Limitations of Traditional Randomized Control Arms
3. External Control Arms as a Fit-for-Purpose Alternative
4. Types of External Control Arms and their Appropriate Use
5. Data Quality Challenges and AI-Based Mitigation in External Control
6. Advanced Statistical and AI Techniques
7. GenAI-enabled External Control Arm (ECA): Rare Disease – Use Case
8. Regulatory and Future Implications

Disclaimer

All opinions expressed during this session are solely those of the speaker and not of current - **Tata Consultancy Services** or past employer. This content is provided for educational and informational purposes only and should not be considered an official stance of my employer.

Emerging Trends in Data-Driven Clinical Trial

Real world evidence
Enables more efficient and patient-centric trial design and driving Insights before, during and after the trials

Therapeutically minded AI, NLP Accelerate real-world data curation and feature extraction

Modern Trial Designs Adaptive, Basket, Umbrella etc.
bringing flexibility & efficiency, improving success rates and being more patient focused

Clinical Data science Clinical Data Science driven Mindset, Insight , Strategies etc.

In Silico Trials, Digital Twins simulate real world trials to reduce, refine and replace trials

Clinical trials are changing. Evidence generation is evolving

Limitations of Traditional Randomized Control Arms

Limited Representativeness

Highly controlled trial criteria and trial settings may lead to under representation of real-world patient diversity, limiting the external validity of trial results.

Ethical Constraints

Use of placebo or withholding active therapy may be ethically challenging for TAs like Rare Diseases , Oncology etc.

Feasibility Constraints

In rare diseases and late-line oncology, limited patient availability can make concurrent control arms impractical.

Operational Burden and Timelines

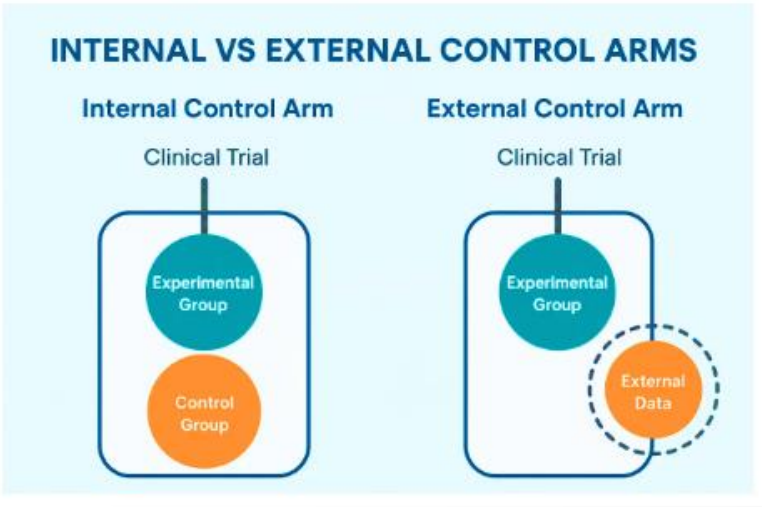
In some programs, constructing and maintaining concurrent control arms can increase operational complexity and timelines

External Control Arms as a Fit-for-Purpose Alternative

External Control Arm (ECA) : A fit-for-purpose comparator derived from high-quality external data (real-world or historical data) when randomization control arm is infeasible or ethically challenging.

Use of External Control Arms

- Help address feasibility, ethical, and recruitment constraints
- May Reduce incremental patient burden and unnecessary placebo exposure
- Enable using broader external datasets that reflect real-world patient populations.
- Support baseline comparability and bias assessment using AI-enabled matching and Bayesian borrowing
- Particularly relevant in rare diseases, oncology, biomarker-defined and late-line therapies



The Scientific Value of External Control Arms in Clinical Research - Gilmore Health News

Types of External Control Arms

Types and
Appropriate Use

Historical Control Arm

- Derived from prior clinical trials or established historical datasets
- Efficient when data and endpoints are well aligned
- Requires careful evaluation of time-related differences to manage bias

Concurrent External Control

- Control patients treated in parallel but outside the trial
- Reduces time-related bias compared to historical controls

Hybrid / Augmented Control

- Combines Small internal control arm with qualified external data
- Bayesian dynamic borrowing may be applied to incorporate external evidence
- Increasingly adopted in oncology and rare disease programs

Synthetic Control Arms

- Constructed using patient-level data from RWD or prior trials
- Uses propensity score matching / weighting to balance baseline covariates
- Preferred approach when supported by patient-level data and robust matching

Data Quality Challenges and AI-Based Mitigation in External Control

Challenges from Real-World Data Heterogeneity

Data Source Heterogenicity

RWD comes from multiple sources like EHRs, claims, registries, and trials, causing structural and coding differences.

Inconsistency and Missing Data

Clinical variables may be inconsistently recorded or missing, especially in rare diseases with limited data availability.

Bias and Temporal Variability

Confounding, selection bias, and changes in treatment practices over time add complexity to data analysis.

Need for Harmonization and AI

Advanced statistical methods and AI-driven cleaning frameworks are essential to harmonize and preprocess RWD effectively.

AI-Driven Quality Control for Real-World Datasets

AI Detection of Data Issues

AI pipelines automatically detect inconsistencies, missing fields, and improbable clinical sequences in real-world datasets.

Natural Language Processing (NLP)

NLP extracts clinically relevant variables from unstructured physician notes and medical reports, expanding usable data.

Machine Learning Anomaly Detection

ML detects outliers such as implausible lab values, duplicate records, or erroneous timestamps that can skew analysis.

Bias Detection and Data Harmonization

Automated algorithms evaluate demographic bias and harmonize coding schemes to standardize diverse clinical datasets.

Advanced Statistical and AI Techniques

METHOD	PURPOSE	STRENGTHS	LIMITATIONS
Propensity Score Matching	Balance covariates between groups	Simple, interpretable	Requires sufficient overlap
Inverse Probability Weighting (IPW)	Reweight observations	Uses whole dataset	Sensitive to extreme weights
Bayesian Dynamic Borrowing	Incorporate external data adaptively	Flexible, probabilistic	Requires expert priors

GenAI-enabled External Control Arm
Rare Disease – Use Case

connect·share·advance

Objective:

- Develop a regulatory-ready External Control Arm using Real-World Data (RWD) for a rare disease trial.
- Demonstrate how GenAI accelerates RWD identification, curation, and matching.

Context:

- Indication: Rare Disease (rare autosomal recessive disorder affecting XXX metabolism).
- Data sources: EHRs and public RWD repositories (~10,000 screened subjects).
- Challenge: Limited randomized controls and small trial populations.

GenAI-enabled RWD Identification & Curation

Inputs

- Trial Protocol, SAP, Treatment-arm data, EHR/RWD

+

GenAI-enabled Processing

- Identification of 'fit-for-purpose' RWD using ICD-10 codes, biomarkers, and clinical symptoms.
- Automatic exclusion of confounding comorbidities.
- Data standardization, de-duplication, and missing data handling.

Outcome

- 2,593 likely Rare Disease subjects identified.
- Shortlisted ~150 high-confidence patients for investigator review.

Statistical Matching & ECA Validation

Map ECA and Treatment Arm Data

Here are the standardized mean differences (SMD) and p-values for key variables:

Variable	Standardized Mean Difference (SMD)	p-value	Balance Interpretation
Age	-0.30	0.099	Good balance (SMD < 0.1 is ideal)
Serum Var1 (mg/dL)	0.94	1.93e-06	Imbalance (SMD > 0.25)
24-hour Urine Var2 (µg/day)	1.23	6.80e-09	Strong imbalance (SMD > 0.25)

- Age is relatively well-balanced, with a small standardized mean difference.
- Serum Var1 and 24-hour urine Var2 still show imbalance, suggesting additional adjustments may be needed (e.g., exact matching, inverse probability weighting).
- p-values confirm significant differences in Var1 and urine Var2 levels between Rare Disease and external control groups.

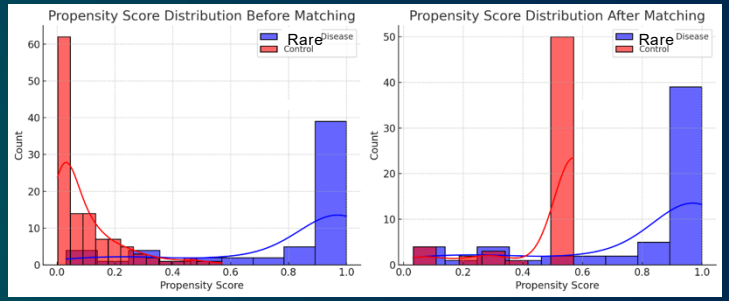
Matching Approach:

- Propensity Score Matching using age, gender, var1, var2, var3, and LFTs.
- Nearest-neighbour matching with balance diagnostics.
- Validation:
- Baseline comparability and balance assessment (SMD, propensity score overlap).
- Bias, confounding, and sensitivity analyses.
- Final Matched Cohort:
- 60 Rare Disease patients and 60 matched external controls.

Contd..

connect·share·advance

Map ECA and Treatment Arm Data



1. **Propensity Score Distribution**
   - The first plot (before matching) shows a clear separation between Rare disease and external control groups.
   - The second plot (after matching) shows an improved overlap, indicating better comparability between groups.

2. **Standardized Mean Differences (SMD) Table**
   - Lower SMD values after matching indicate improved balance for most covariates.
   - **Serum Var1 (µg/dL)** saw a slight increase in imbalance due to synthetic data but remains within an acceptable range.
   - **Serum Var2 (mg/dL) and 24-hour Urine Var3 (µg/day)** still show some imbalance, but it has improved.

Summary of Results for Rare Disease Matched Control Study

Objective:

To evaluate clinical outcomes in Rare disease patients compared with a propensity score–matched external control group.

Study Design:

A matched cohort analysis of 60 Rare disease patients and 60 external controls, using propensity scores based on biochemical and demographic factors.

Matching Criteria:

- Age
- Gender
- Serum Var1
- Serum Var2
- 24-hour Urine Var3

Baseline Characteristics:

Variable	Rare Patients (N=60)	External Controls (N=60)
Age (years)	29.57 ± 12.17	26.67 ± 4.19
Serum Var1 (mg/dL)	8.05 ± 3.92	10.88 ± 1.46
Serum Var2 (µg/dL)	33.41 ± 15.58	37.50 ± 6.63
24-hour Urine Var3 (µg/day)	496.37 ± 44.04	727.59 ± 35.92

Key Results, Regulatory & Business Value

Key Results

- Significant differences in metabolism of chemical markers between ECA and disease cohorts.
- Improved balance post-matching, supporting ECA validity.

Regulatory Value

- Transparent, reproducible ECA validation framework.
- Supports evidence generation in rare diseases with limited controls.

Business Value

- Reduced manual effort and cycle time using GenAI.
- Scalable ECA framework reusable across indications.

AI-Enabled Impact on Clinical Development

Acceleration of Clinical Development Timelines

Reduction in Patient Recruitment

AI-driven external control arms eliminate the need for placebo groups, reducing patient recruitment by **30–50%.**

Streamlined Data Preparation

AI tools cut data curation from months of manual work to days or weeks, improving trial efficiency.

Faster Trial Decisions

Machine learning enables rapid patient matching and interim analyses, accelerating adaptive trial designs.

Cost and Time Savings

AI-enhanced ECAs reduce trial sites and monitoring needs, saving costs and shortening timelines by **30–40%.**

Regulatory and Future Implications

Regulatory Recognition of ECAs

Global regulatory bodies support External Control Arms enhanced by AI for rare diseases and oncology trials when fit-for- purpose

AI Enhancing Regulatory Confidence

AI enables transparent data curation, baseline matching, and quality control to meet audit and reproducibility standards.

Future Clinical Trial Innovations

ECAs will support adaptive trials, federated learning, and integration of multimodal data for patient-centric evidence.

connect·share·advance

THANK YOU!