Architecting Trust: Making Real-World Data Submission-Ready for Rare Disease and Oncology Development

Biographies

Anbu Damodaran

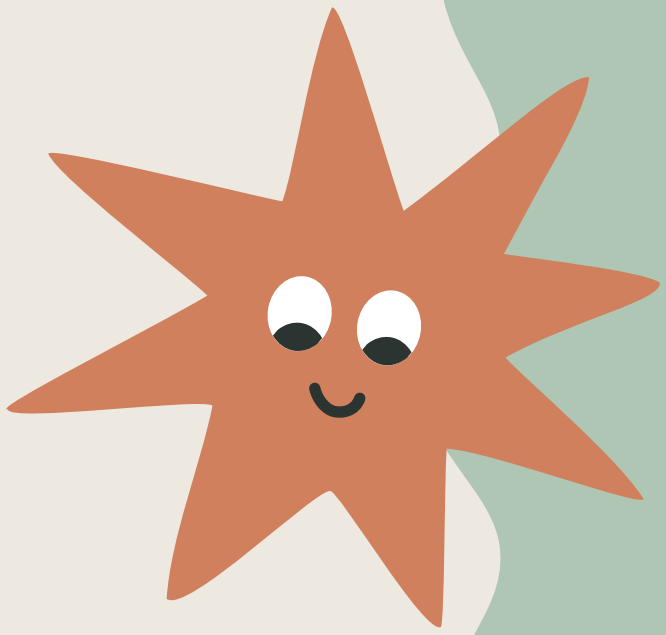
Anbu Damodaran is a seasoned drug development professional with over 25 years of experience in biometrics and clinical data operations, supporting studies ranging from first-in-human trials to post-market commitments. He has a deep understanding of the end-to-end drug development process, including regulatory submissions and compliance with ICH, GCP, CFR, and FDA guidelines. Anbu leverages his expertise in risk and change management while applying emerging AI trends to modernize biometrics workflows, particularly in rare-disease development. He currently serves as Associate Director of Statistical Programming at Alexion, AstraZeneca Rare Disease.

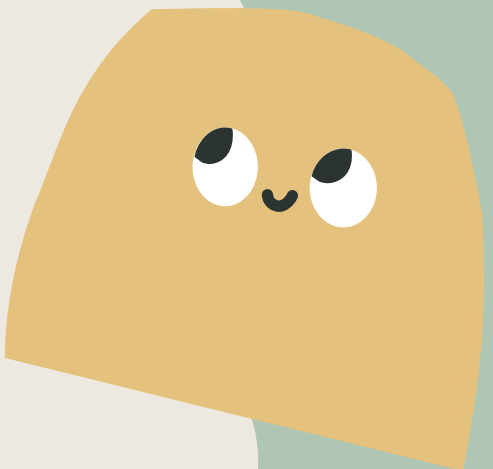
Satya Medapati

Satyananda (Satya) Medapati is an accomplished biopharmaceutical leader with over 25 years of experience in statistical programming, clinical data, and biometrics operations. He has led global teams across multiple geographies, delivering high-quality analytics and regulatory-ready outputs across the full drug development lifecycle, from early clinical development to post-marketing commitments. Known for driving operational excellence, automation, and compliance with global regulatory standards, he currently serves as Director of Statistical Programming at Alexion, AstraZeneca Rare Disease in Bangalore, India.

Architecting Trust: Making Real-World Data Submission-Ready for Rare Disease and Oncology Development

Anbu Damodaran & Satya Medapati





ARCHITECTING TRUST: MAKING REAL-WORLD DATA SUBMISSION-READY FOR RARE DISEASE AND ONCOLOGY DEVELOPMENT

Abstract: This presentation explores transforming Real World Data (RWD) into regulatory grade evidence for rare disease and oncology development. We examine the evolution from legacy historical controls to modern External Control Arms (ECAs) using sophisticated methodologies such as Propensity Score Matching and Indirect Treatment Comparisons. To bridge the regulatory trust gap, the session details a framework for semantic harmonization, clinical grade quality control, and data traceability. We highlight the critical role of statistical programmers in navigating complex data mapping challenges, such as RWD to ADaM transitions, while maintaining CDISC alignment. By establishing a robust Trust Architecture through transparent documentation and rigorous causal inference, researchers can leverage RWD to overcome recruitment barriers and accelerate the delivery of innovative therapies.

DISCLAIMER

This presentation is intended solely for informational purposes and should not be interpreted as professional advice or guidance. The views and opinions expressed are those of the presenter and do not necessarily represent the views or policies of AstraZeneca.

ARCHITECTING TRUST: MAKING REAL-WORLD DATA SUBMISSION-READY FOR RARE DISEASE AND ONCOLOGY DEVELOPMENT

Agenda

- Introduction to Real-World Data
- Comparative Evidence Methodologies
- RWD External Controls
- Regulatory Perspectives
- Programming Considerations (SDTM, ADaM, TLFs)
- Documentation & Reviewer Expectations
- Future Directions

Real-World Data Foundations

WHY RWD IN RARE DISEASE & ONCOLOGY

Challenges of Traditional Trials

RCTs in rare diseases face small populations, ethical issues, and urgency to accelerate therapies.

RWD Enables External Controls

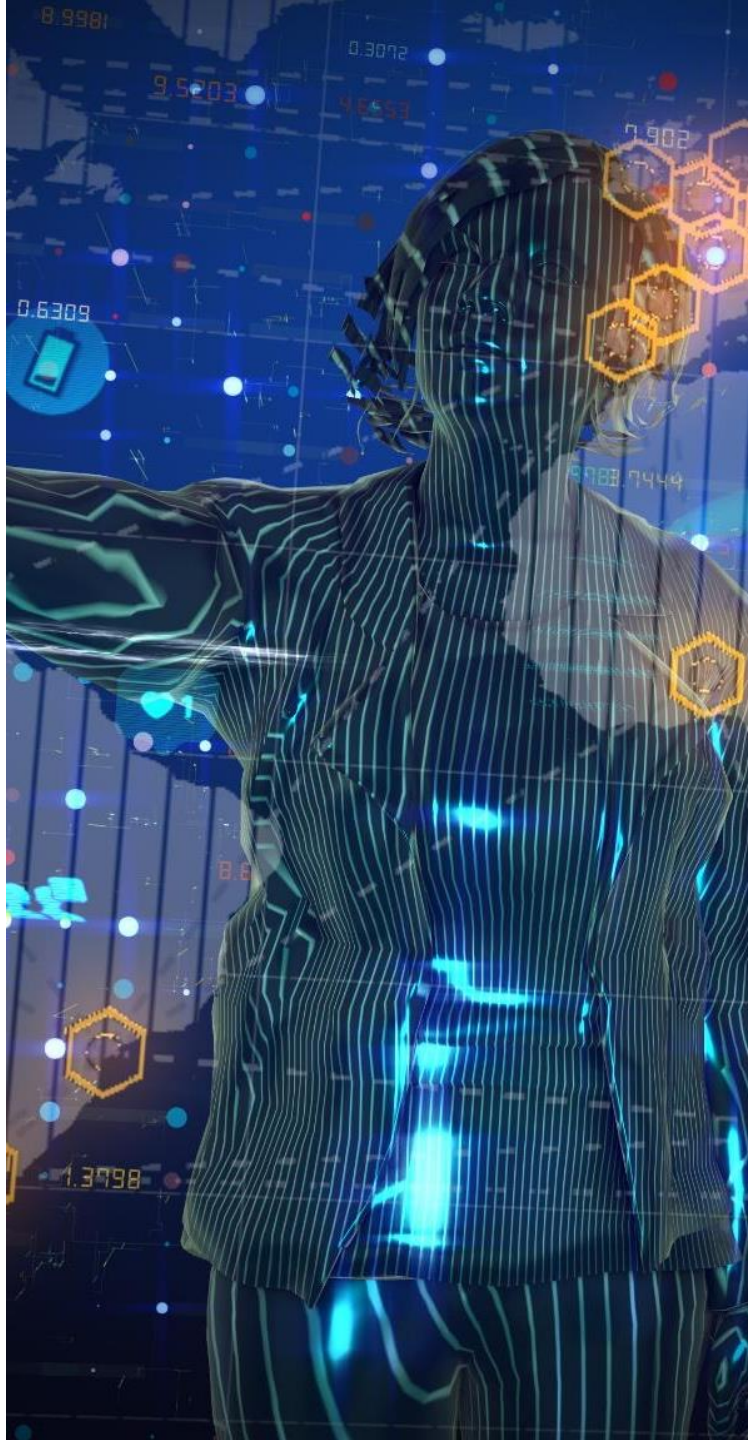
RWD allows creation of external control arms for single-arm trials, contextualizing treatment effects without randomization.

Broader Patient Representation

RWD captures diverse patient populations and real clinical treatment patterns improving study relevance.

Addressing Data Challenges

Developers must manage data quality, selection bias, and heterogeneity to ensure RWD reliability and regulatory acceptance.



Overview of Real-World Data in Clinical Research

Definition and Sources of RWD

Real-world data comes from electronic health records, registries, medical claims, and digital health technologies.

RWD in Drug Development

RWD supports rare disease and oncology research by enabling more inclusive evidence beyond clinical trials.

Regulatory and Analytical Advances

Regulatory agencies encourage robust RWD use with frameworks for causal interpretation and high-quality evidence.

Challenges and Importance

Understanding data provenance, completeness, and relevance is crucial for integrating RWD in research and submissions.

Comparative Evidence Methodologies

Anchored ITC, NMA, MAIC, Historical Controls, and External Controls

Anchored Indirect Treatment Comparisons (ITCs)

ITCs compare two treatments sharing a common comparator, maintaining internal validity across separate randomized trials.

Network Meta-Analysis (NMA)

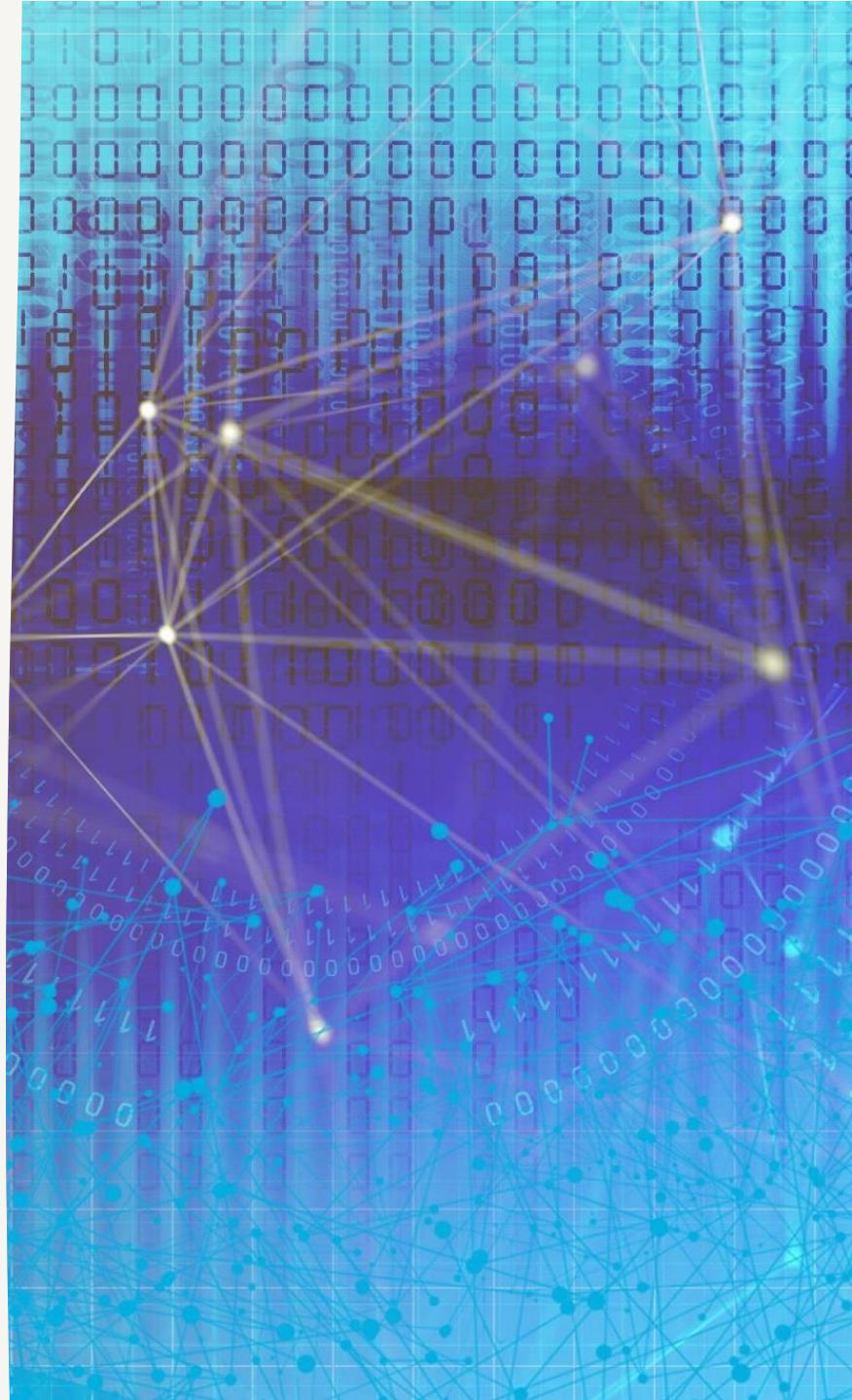
NMA estimates relative effects across multiple treatments simultaneously, assuming consistency throughout the evidence network.

Matching-Adjusted Indirect Comparisons (MAICs)

MAICs adjust for population differences by reweighting patient data to compare single-arm studies with randomized trials.

Historical and External Controls

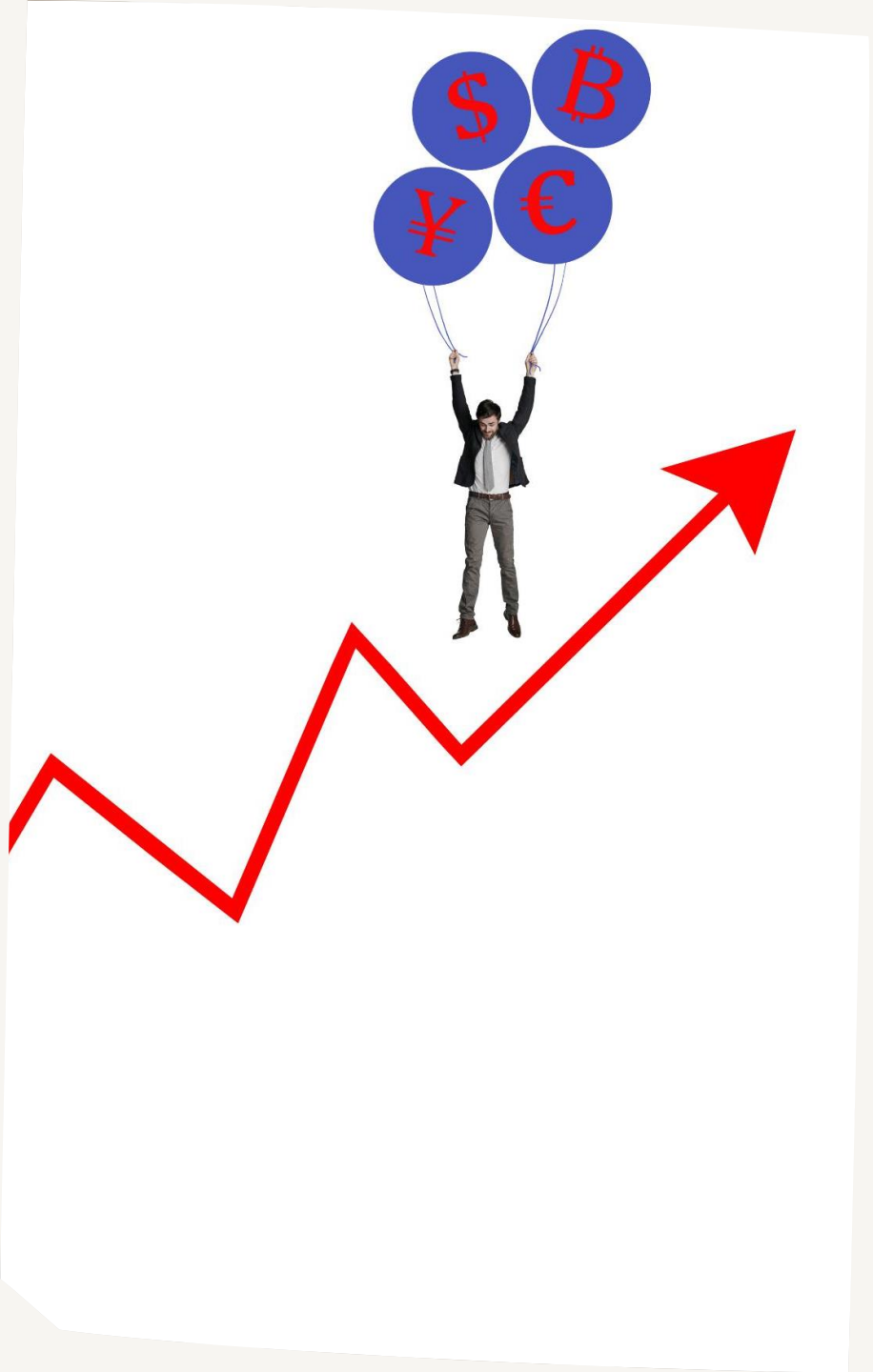
Historical controls use past patient data, while externally controlled trials incorporate real-world data to support rigorous comparisons.



The "Trust" Gap – Comparing Methods

This table illustrates the trade-offs between connectivity requirements and potential bias:

Method	Requires Common Comparator?	Bias Level	Typical Use Case
Anchored ITC/NMA	Yes	Low–Moderate	HTA submissions
MAIC	No	Moderate–High	Payer submissions
Historical Control	No	Very High	Ultra-rare diseases
External Control Arm	No	Moderate–High	Rare disease/Oncology



Biases and Mitigation Strategies

Confounding Bias Management

Address confounding via propensity-score modeling, weighting, stratification, or matching to balance covariates.

Selection Bias Reduction

Mitigate selection bias by harmonizing eligibility criteria to mirror trial inclusion and exclusion standards.

Temporal Bias Control

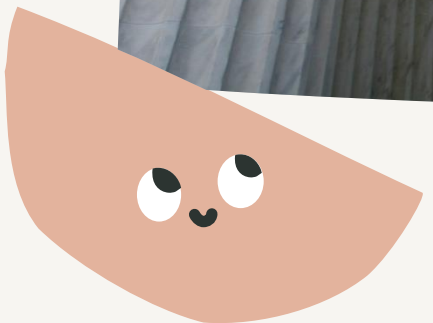
Minimize temporal bias using contemporaneous data or stratifying analysis by calendar periods.

Immortal Time Bias Correction

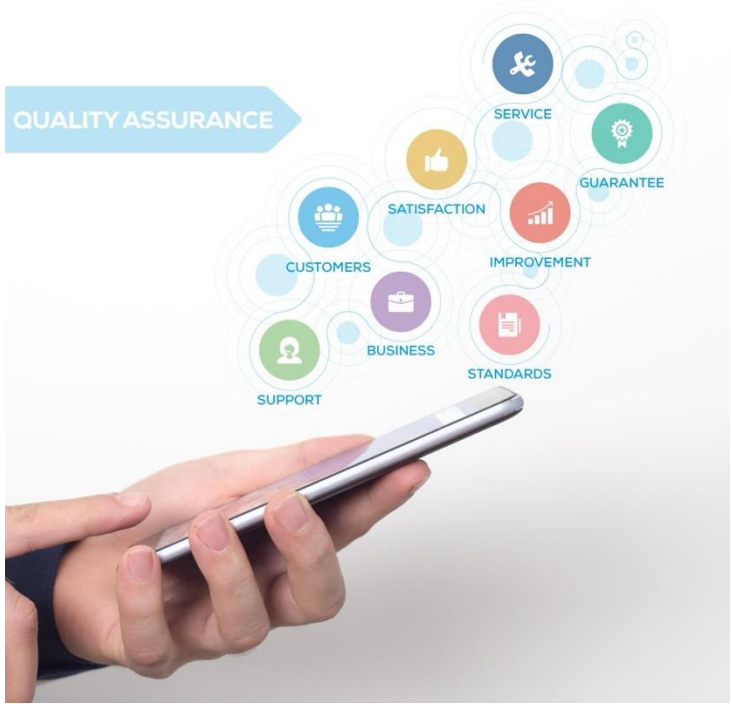
Define explicit index dates and remove immortal periods to avoid artificially extended survival outcomes.

The Trust Architecture: Four Pillars of Credibility

- **Pillar 1:** Semantic Harmonization and Endpoint Fidelity—aligning protocol endpoints to heterogeneous source systems.
- **Pillar 2:** Automated Quality Control—scalable checks for clinical-grade standards in non-trial environments.
- **Pillar 3:** Principled Causal Inference—emulating target trials to neutralize bias.
- **Pillar 4:** CDISC-Aligned Statistical Programming—reproducible workflows for SDTM and ADaM.



CHRONOLOGY & GUIDANCE



Foundational Regulatory Frameworks

ICH E10 (2000/2001) defined external controls and control-arm selection principles, setting the groundwork for comparative evidence methods.

Adoption of Advanced Methods

During the 2010s, HTA bodies and regulators adopted Network Meta-Analysis and Matching-Adjusted Indirect Comparisons for value assessments.

Recent Regulatory Milestones

Key recent milestones include FDA's 2023 draft guidance, EMA's 2025 reflection paper, and PMDA's RWD-based external control use in orphan drugs.

Regulatory Convergence with Regional Nuances

While global regulatory convergence is increasing, region-specific nuances remain crucial when preparing regulatory submissions.

THE EVOLUTION OF EVIDENCE SYNTHESIS

1950s–1980s

Evidence relied heavily on historical controls, which were often simple, pragmatic, but highly biased.

1990s

The introduction of the Bucher methodology (~1997) allowed for anchored indirect comparisons using shared trial arms.

Early 2000s

Network Meta-Analysis (NMA) became the standard for Health Technology Assessment (HTA), enabling comparisons across full treatment networks.

2010s

The development of Matching-Adjusted Indirect Comparison (MAIC) allowed researchers to address heterogeneity in trials without shared comparators.

2015–Present

The rise of modern External Control Arms (ECA) reflects a shift toward using robust Real-World Data (RWD) infrastructure for regulatory submissions.

WHEN ECAS ARE APPROPRIATE

Ethical Use of ECAs

ECAs are suitable when randomization is unethical or infeasible, especially in aggressive or ultra-rare conditions.

Scientific Rationale

ECAs are justified when effect sizes are large, natural history is well-understood, or long-term outcomes require real-world data.

Feasibility and Bias Mitigation

Feasibility and rigorous design must guide ECA adoption to minimize bias and ensure valid results.

Concurrent vs Historical Controls

Concurrent ECAs are preferred, but well-documented historical controls can be acceptable with proper methodology.

DESIGNING AN ECA: TARGET-TRIAL EMULATION

Defining Study Framework

Establish clear time-zero, align eligibility criteria, and specify treatment strategies for trial and external cohorts.

Endpoint and Confounder Alignment

Choose consistent endpoints and pre-specify confounders and effect modifiers to ensure valid comparisons.

Sensitivity Analyses Planning

Plan sensitivity analyses to assess immortal-time bias, calendar-time effects, and impacts of missing data.

Enhancing Causal Interpretability

Rigorous target-trial emulation improves causal inference and boosts regulatory confidence in evidence.

RWD as an External Control

Using RWD to Construct External Control Arms

Data Source Identification

Select high-quality data sources like disease registries and electronic health records with detailed exposure and outcome data.

Alignment of Study Criteria

Ensure eligibility criteria, baseline characteristics, and endpoints closely match the investigational arm for valid comparisons.

Causal Inference Techniques

Apply methods like propensity score matching and inverse probability weighting to reduce confounding in external control arms.

Regulatory Transparency and Sensitivity

Incorporate sensitivity analyses and address confounding, misclassification, and temporal mismatches to meet regulatory standards.

Data Sources & Provenance

Varied Real World Data Sources

Includes EHRs, registries, claims, lab feeds, radiology, and digital health technologies in regulatory data landscape.

Data Linking and Quality Challenges

Challenges include linking, deduplicating, and qualifying heterogeneous data for reliable regulatory use.

Documentation and Transparency

Document extraction windows, refresh cycles, and data lineage to ensure transparency and reliability.

Assessing Data Validity and Bias

Evaluate completeness, measurement validity, and potential biases to maintain credible data for ECAs.

Confounding Control Toolkit

Propensity Models and Weighting

Use logistic and machine-learning propensity models with IPTW to control confounding in real-world data analyses.

Matching and Stratification

Matching subjects and stratifying data help balance covariates and reduce bias in comparative studies.

MAIC Weighting and Diagnostics

MAIC weighting adjusts for baseline imbalances; balance diagnostics and ESS ensure data validity and overlap.

Sensitivity Analyses

Negative controls and quantitative bias analyses test robustness and help confirm causal inferences.

Global Regulatory Perspectives

FDA, EMA, and PMDA Approaches to RWD External Controls

FDA's RWE Framework

The FDA emphasizes fit-for-purpose data quality, prespecified analysis plans, and transparent documentation in its Real-World Evidence framework.

EMA's Evolving Guidance

EMA highlights challenges in consistency, data interoperability, and methodological quality while developing guidance on external controls from RWD.

PMDA's RWD Use Cases

PMDA focuses on RWD use in orphan disease trials, stressing data reliability, planning, and careful interpretation for regulatory decisions.

Global Regulatory Consensus

All agencies recognize RWD's potential to accelerate evidence generation while insisting on rigorous methods to reduce bias.

Regulatory Views: FDA

- Finalized Frameworks (2024–2025)
- Finalized guidance for EHR/Claims (July '24) and Devices (Dec '25) establishes rigorous 'Fit-for-Purpose' standards.
- De-identified Data Access (Dec 2025)
- Major policy shift: FDA now accepts de-identified datasets (e.g., SEER, claims) without requiring identifiable patient-level data for specific submissions.
- 2026 Traceability & AI Priorities
- New focus on end-to-end traceability from source to submission and AI-enabled RWD analysis for oncology endpoints.

Regulatory Views: EMA

EMA's Interest in External Controls

EMA shows growing interest in external controls highlighted in its 2025 concept paper guiding future regulatory reflections.

Use Cases Supporting ECAs

Use cases like CAR-T therapies demonstrate that well-designed ECAs can support centralized regulatory procedures effectively.

Key Regulatory Expectations

EMA expects clarity on definitions, minimal data needs, feasibility, and bias-control strategies in ECA submissions.

Sponsor Preparation Requirements

Sponsors must prepare clear rationales and transparent methodologies when submitting ECA-based evidence to EMA.

Regulatory Views: PMDA

Full-Lifecycle Planning

PMDA emphasizes planning studies from design to interpretation ensuring comprehensive regulatory review.

Data Reliability and Transparency

Reliable, transparent data is critical for regulatory acceptance of RWD to support approvals.

Methodological Rigor

Careful data curation and defensible methodologies underpin PMDA's principles for external control use.

Regulatory Precedents

Drugs Approved Using RWD as External Control Evidence

Role of Real-World Data

Real-world data is used as external or supplemental evidence to support drug approvals when randomized trials are challenging.

Examples of Drug Approvals

Drugs like Nulibry, Ibrance, and Prograf used real-world data to support approval and/or label expansion in rare or complex cases.

Regulatory Considerations

Regulators require well-curated data, minimized confounding, and transparent analytics for real-world evidence acceptance.

Importance of Data Integrity

Transparent data processing and reproducible analytics are crucial for justifying non-randomized comparator use.

Patient Linkage Approaches and Evaluation in RWD

Record Linkage and Tokenization Methods

Deterministic Record Linkage

Deterministic linkage matches exact or normalized identifiers like name tokens and date of birth for accurate data integration.

Probabilistic Record Linkage

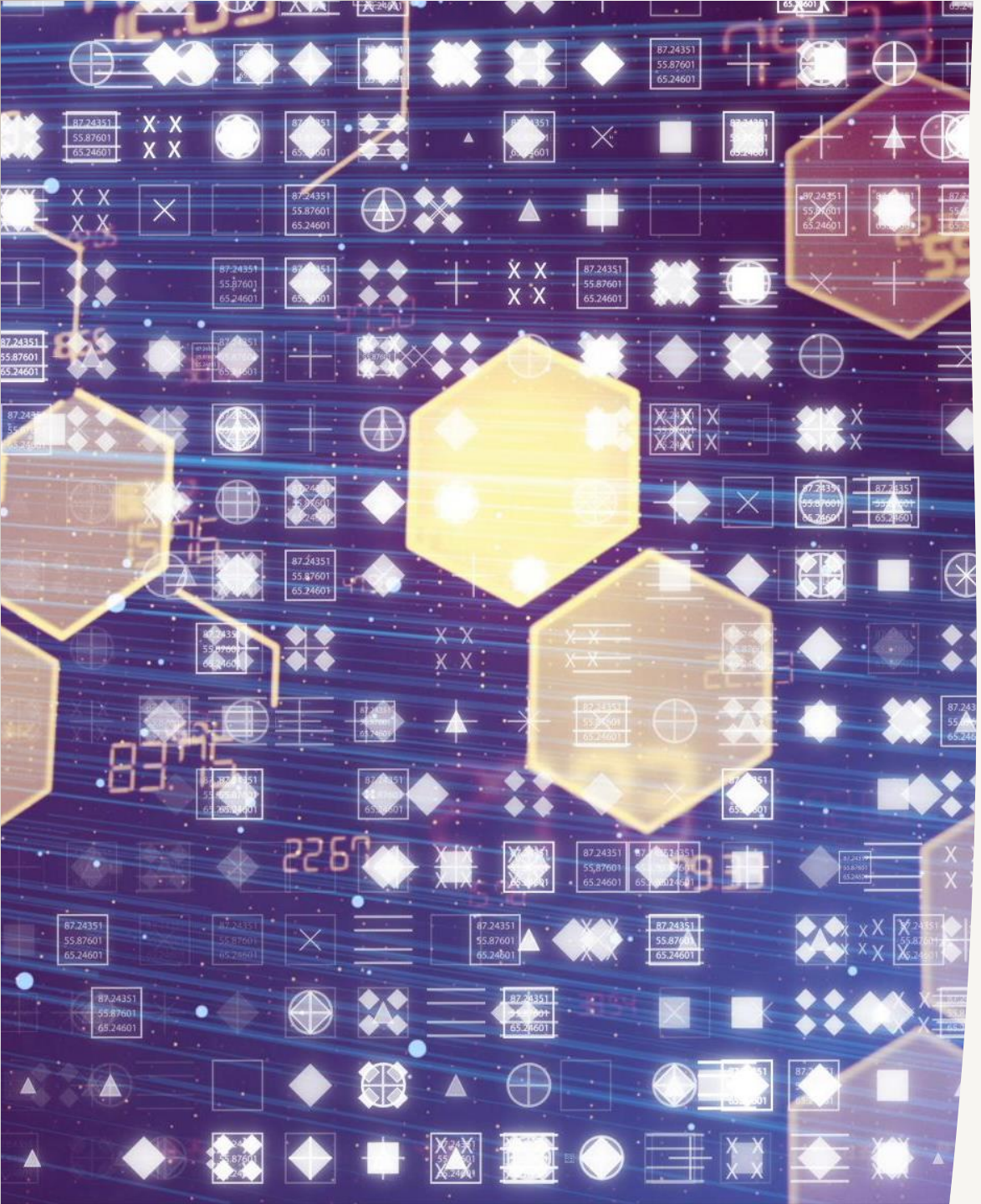
Probabilistic linkage uses similarity metrics and statistical models to match imperfect or incomplete real-world data.

Privacy-Preserving Linkage

PPRL methods secure data privacy by using hashing, cryptographic transforms, and Bloom filters for confidential matching.

Tokenization Techniques

Tokenization creates irreversible, consistent identifiers enabling secure data linking without exposing real identifiers.



Advanced ML-Based Matching, Overlap Assessment, and Tools

Machine Learning Approaches

Supervised and unsupervised ML methods enhance patient-level data matching using classifiers and clustering techniques.

Overlap and Concordance Assessment

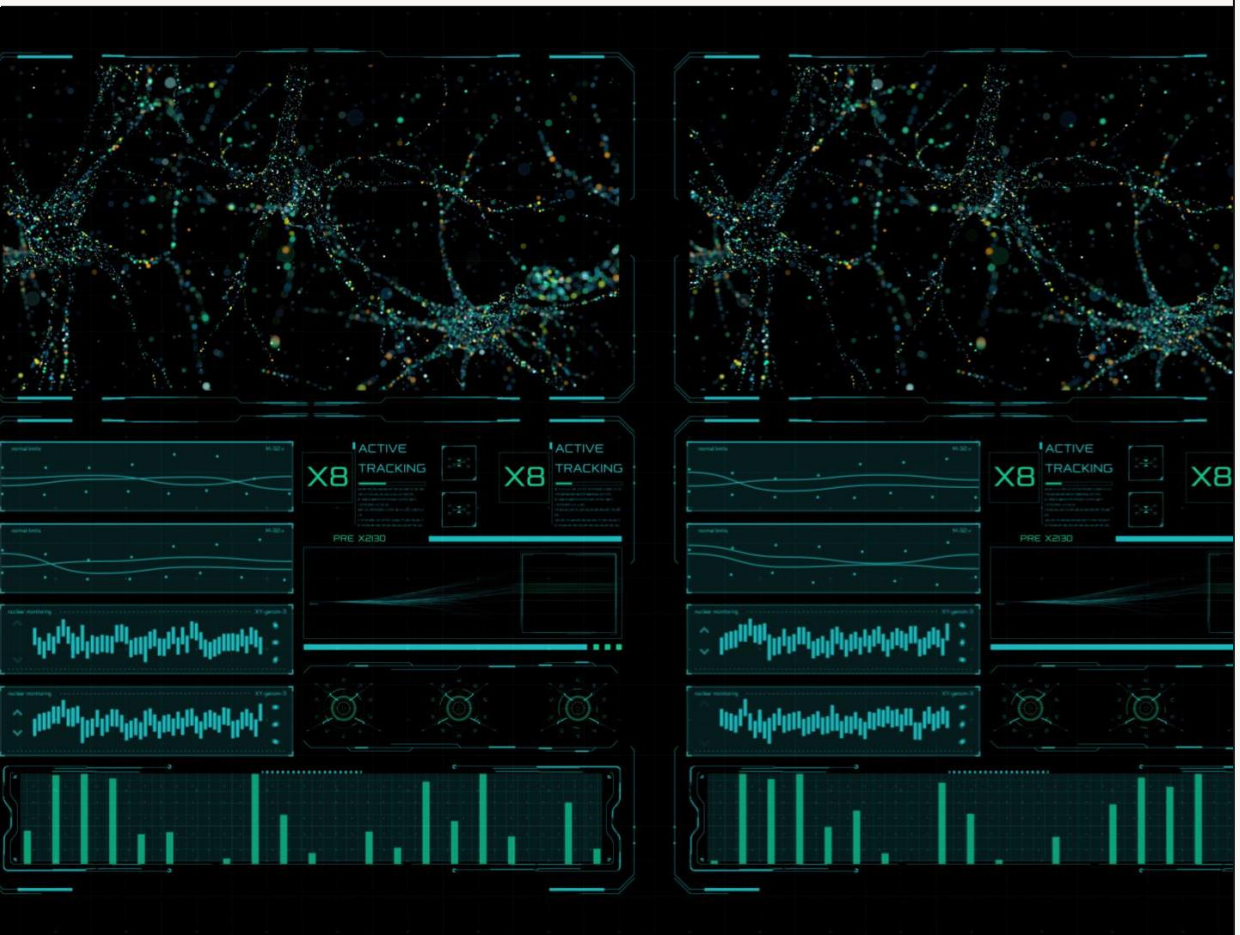
Metrics quantify data overlaps and discrepancies, using visualizations like Venn diagrams, scatter plots, and timeline alignments.

Toolkits for Data Linkage

Tools like Splink, Febrl, and Dedupe support scalable and customizable RWD linkage and enrichment workflows.

Programming Considerations: SDTM

SDTM MAPPING CHALLENGES WITH RWD



RWD Complexity in SDTM Mapping

Real-world data lacks standardization, causing challenges in consistent mapping to SDTM domains.

Data Harmonization Challenges

Laboratory data units vary greatly and require harmonization and terminology alignment like LOINC.

Handling Missing or Inferred Data

Exposure data may be missing in EHRs, requiring inference from related medication or claims data.

Ensuring SDTM Compliance and Transparency

Accurate mapping with detailed annotations ensures transparency and data fidelity in SDTM datasets.

How External Controls Affect SDTM Mapping

Handling Data Heterogeneity

Reconcile diverse coding systems and terminologies into SDTM-compliant formats for consistent data integration.

Deriving SDTM Domains

Select variables, establish consistent start dates, and harmonize visit timings despite irregular real-world data.

Algorithmic Endpoint Inference

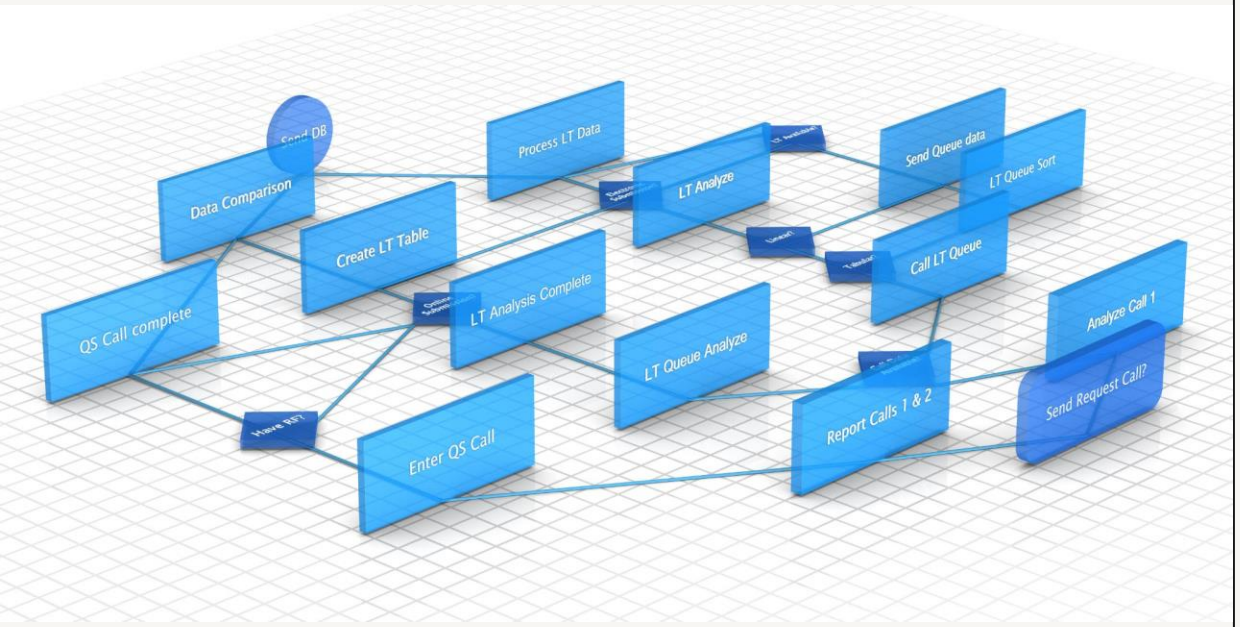
Use rule-based logic to infer progression dates, treatment exposures, and adverse events from non-standardized data.

Provenance and Documentation

Track data provenance and flag missing data to ensure transparency and reproducibility in regulatory review.

Programming Considerations: ADaM

DIRECT-TO-ADAM MAPPING WHEN SDTM IS NOT FEASIBLE



Core Challenges:

Lack of Standard Terminology in source RWD

High Complexity in mapping unstructured ‘messy’ data to SDTM domains

Risk of losing critical data granularity during multi-step transformations

The ‘Direct-to-ADaM’ Approach:

Bypasses SDTM to maintain analytical ‘line-of-sight’ from source to endpoint

Focuses on Traceability through robust Analysis Data Reviewers Guide (ADRG)

Regulatory Mitigation: Early Alignment

Type-C meeting: Highly recommended to secure formal FDA agreement prior to submission

Objective: Concurrence on data architecture, rationale for SDTM bypass, and traceability plan

Documentation: Reference meeting minutes in the CSR and ADRG

Impact on ADaM Datasets and Causal Modeling

Complexity with Real-World Data

Integrating real-world data requires additional variables and baseline covariates for causal inference techniques in ADaM datasets.

Causal Inference Metrics

Effective sample size and covariate balance diagnostics evaluate adjustment quality in causal modeling.

Time-to-Event Dataset Harmonization

Creating ADTTE datasets involves harmonizing censoring rules and performing sensitivity analyses for inconsistent follow-up.

Transparency and Documentation

Clear definitions of treatment strategies, eligibility, and time-zero anchors ensure regulatory-grade causal evidence.

Programming Considerations: Documentation

cSDRG, ADRG, and Reviewer Documentation Requirements

cSDRG Documentation Details

The cSDRG describes data sources, extraction, linkage, deduplication, transformations, and outlines data quality and limitations.

ADRG Analytical Dataset Construction

ADRG explains dataset construction, including causal models, weighting, confounder selection, and sensitivity analyses.

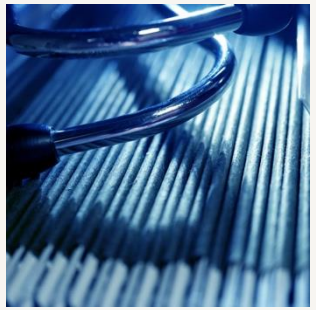
Traceability and Regulatory Expectations

Traceability between SDTM and ADaM is crucial, with transparent assumptions to infer clinical events for regulatory review.

Supporting Appendices and Credibility

Appendices include metadata, programming flowcharts, and cohort attrition diagrams to ensure reproducibility and credibility.

Precedents & Case Signals

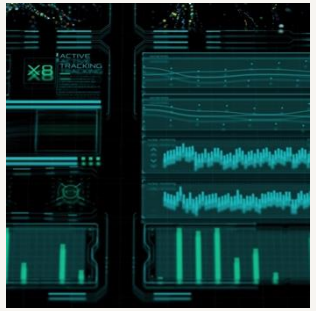


Regulatory Approvals with RWD

Real-world data and external controls have supported regulatory approvals, particularly in hematology and oncology.

CAR-T and Single-Arm Programs

CAR-T therapies and single-arm clinical trials utilized external control benchmarks for evaluation and approval.



Key Lessons for ECA Use

Replicated methodology, transparent reporting, and citing regulatory precedents strengthen external control arm justification.

Conclusion and Future Directions

Key Takeaways and the Future of RWD in Trials

Innovative Trial Designs

Real-world data enables innovative trial designs reducing reliance on traditional control arms in rare diseases and oncology.

Role of Statistical Programmers

Statistical programmers ensure data quality and regulatory-grade evidence through robust analyses and documentation.

Future Opportunities

Future RWD use includes hybrid designs, machine learning models, real-time data linkage, and decentralized trials.

Ethical and Scientific Integrity

Proper use of RWD requires diligence, clarity, and scientific rigor to broaden therapy access and reduce patient burden.

What's Next

Hybrid ECAs Integration

Future hybrid ECAs combine randomized and external data to improve evidence robustness in clinical research.

Anchored ML-NMR Techniques

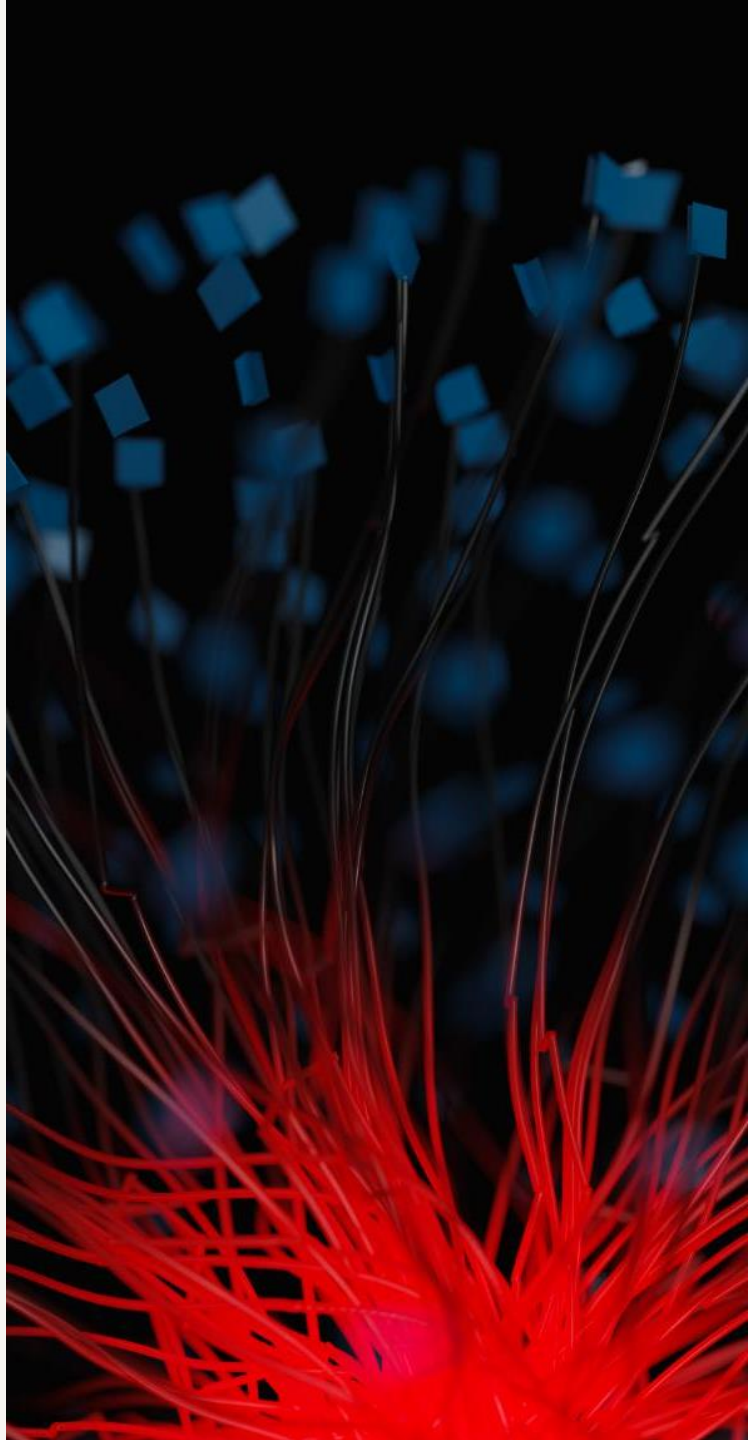
Anchored machine learning NMR methods enhance network analysis accuracy in medical research.

Tokenization for Data Linking

Expanding tokenization links clinical trials with real-world datasets for longitudinal evidence generation.

Decentralized Trials and Digital Phenotyping

Decentralized trial designs and digital phenotyping broaden real-world data availability and insights.



QUESTIONS?

Name: Anbu Damodaran

Affiliation: Alexion, AstraZeneca Rare Disease

E-mail: anbu.damodaran@alexion.com

REFERENCES

- Liu F. Data Science Methods for Real-World Evidence Generation in Real-World Data. Annu Rev Biomed Data Sci. 2024;7:201-224.
- Zhu R, et al. Clinical pharmacology applications of real-world data and real-world evidence in drug development and approval—an industry perspective. *Clin Pharmacol Ther*. 2023;114(5):973-987.
- Precision for Medicine. Rare neurological disease trials: smarter design with natural history, external controls, and real-world evidence. *Clinical Leader*. August 20, 2024.
- Lin L, Lucassen MJJ, van der Noort V, Egberts TCG, Beijnen JH, Huitema ADR. The feasibility of using real world data as external control arms in oncology trials. Drug Discov Today. 2025 Mar;30(3):104324.
- Clinical Leader. How an external control arm changed phase 3 trials for brain cancer. *Clinical Leader*. August 19, 2021.
- Jaksa, A., Louder, A., Maksymiuk, C., Vondeling, G. T., Martin, L., Gatto, N., ... & Rosenlund, M. (2022). A comparison of 7 oncology external control arm case studies: critiques from regulatory and health technology assessment agencies. *Value in Health*, *25*(12), 1967-1976.
- Yap, Timothy A., et al. "Application of Real-World Data to External Control Groups in Oncology Clinical Trial Drug Development." *Frontiers in Oncology*, vol. 11, 2022, 695936.
- Burger RJ, Ho MP, Woodworth TS, et al. Use of real-world data for external control arms in regulatory decision-making: a systematic review. *Value Health.* 2021;24(9):1367-1374.
- Schmidt M, Gertz R, Radzikh I, et al. External control arms in oncology: a review of recent regulatory submissions. *Ther Innov Regul Sci.* 2024.
- Zou, K.H., Vigna, C., Talwai, A. *et al.* The Next Horizon of Drug Development: External Control Arms and Innovative Tools to Enrich Clinical Trial Data. *Ther Innov Regul Sci* **58**, 443–455 (2024).
- U.S. Food and Drug Administration. Use of Real-World Evidence to Support Regulatory Decision-Making for Medical Devices: Guidance for Industry. Silver Spring, MD: FDA; 2025.
- U.S. Food and Drug Administration. Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products: Guidance for Industry. Silver Spring, MD: FDA; 2024.
- U.S. Food and Drug Administration. Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products: Draft Guidance for Industry. Silver Spring, MD: FDA; 2024.
- European Medicines Agency. Reflection paper on use of real-world data in non-interventional studies to generate real-world evidence. Amsterdam, Netherlands: EMA; 2025. EMA/CHMP/402275/2023.
- European Medicines Agency. Good Practice Guide for the use of the HMA-EMA Catalogues of real-world data sources and studies. Amsterdam, Netherlands: EMA; 2025. EMA/131019/2023.
- Pharmaceuticals and Medical Devices Agency. Points to consider for externally controlled trials (Early Consideration). Tokyo, Japan: PMDA; 2025.
- Pharmaceuticals and Medical Devices Agency. Basic principles on utilization of registry for applications. Tokyo, Japan: PMDA; 2021.