The Curious Case of External Controlled Arms (ECA): Practical Solutions for External and RWD Integration

Gautham Selvaraj, Angelo Tinazzi
PBS Programming , Cytel

Meet the Speakers

Gautham Selvaraj

Title: Associate Director Stat programming

Organization: Cytel Inc

Having 17 years of experience in clinical statistical programming, with expertise in end-to-end clinical data processing aligned with CDISC and sponsor-specific standards. Have demonstrated proficiency in eCTD package submissions across multiple therapeutic areas, including oncology, diabetes, neuroscience, and immunology.

Outside of work, enjoy playing badminton during leisure time.

Angelo Tinazzi

Title: Senior Director Stat Programming

Organization: Cytel Inc

30 years of experience across Italy, the UK, and Switzerland, Angelo leads data standards initiatives at Cytel, advising clients and internal teams on best practices for regulatory submissions to health authorities. He also supports application development and automation initiatives within Cytel's PBS Statistical Programming Group. Additionally, he authors the Cytel Good Data Submission Doctor blog series.

Angelo is a CDISC ADaM Authorized Instructor and a member of the CDISC European Coordinating Committee, where he leads the Italian-speaking User Network.

Disclaimer and Disclosures

- *The views and opinions expressed in this presentation are those of the author(s) and do not necessarily reflect the official policy or position of CDISC.*

- *The author(s) have no real or apparent conflicts of interest to report.*
- *All sensitive information has been anonymized to maintain confidentiality. Our intention is to share our experience in conducting ECA studies.*

Agenda

Why ECA is important

Challenges in incorporating ECA and CDISC package

FDA's View

Our Case studies

Learnings for future considerations

Why ECA is important?

Trial Conduct challenges

- In Rare Disease, Recruitment of Subjects in Randomized control trials is challenging, time-consuming, resource intensive and expensive.
- It is also unethical to administer Placebo Arms in life threatening disease or Conditions.

Existing data for ECA

- Instead, we utilize historical clinical data, natural history studies and Real-world data to supplement or replace control arms.
- RWD may represent more diverse and realistic patient populations, improving external validity of trial results.

Accelerating drug development

- Reduce the need of large control groups which can accelerate trial timelines and regulatory submissions.

Additional usage

- Apart from NDA, this can be used in comparative effective research, Post Marketing studies, label expansion where a randomized trial is not required or feasible.

Challenges in incorporating ECA

Data Quality & Diversity

- Diverse data sources, inconsistent data collection methods
- Missing or incomplete data, Misaligned variable names, formats
- Endpoint definitions and timing often vary across sources

CDISC package

- Ensuring clear traceability from raw to analysis datasets is difficult with legacy data
- External data often lacks standard structure this requires significant transformation to CDISC formats.
- Real-world datasets often lack detailed metadata (e.g., CRF annotations), complicating define.xml creation

Selection Bias

- No randomization baseline imbalances (e.g., severity, age, comorbities)
- Channeling bias: treatment decisions influenced by prognosis
- Requires advanced methods (e.g., PSM, IPW) to reduce confounding

Programmatic Challenges

- Data integration requires cleaning, mapping, and CDISC alignment
- External and trial data must be harmonized to enable integrated analysis and submission-ready datasets
- eCTD package preparation must be precise and compliant

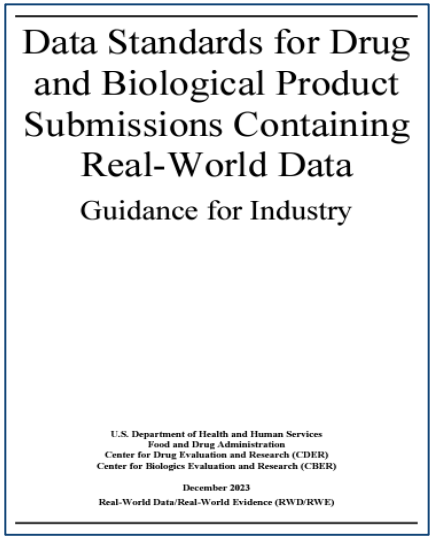
FDA view on ECA

Current guidance

- *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products* (FDA, 2023)
- *Data Standards for Drug and Biological Product Submissions Containing Real-World Data* (FDA, 2024)

When ECA's are appropriate
Emphasize strict requirements for data reliability and validity.
Need for high quality, well documented RWD sources.
Well defined and consistent endpoints.
Importance of bias mitigation.
Alignment with CDISC Standards.
The FDA is open-minded but cautious: ECAs are not a shortcut to approval.

Current Guidance and Industry Initiatives


Data Standards for Drug and Biological Product Submissions Containing Real-World Data

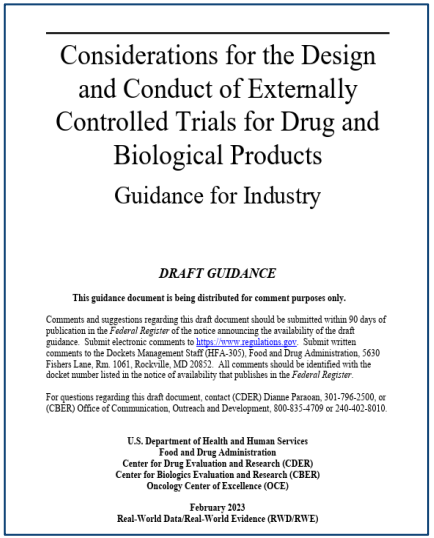
Guidance for Industry

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

December 2023

Real-World Data/Real-World Evidence (RWD/RWE)

- **RWE as whole**
- Focus on **SDTM** only
- Recommendation for **define.xml** and **csdrg**
- Differences in coding system


Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products

Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to https://www.regulations.gov. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

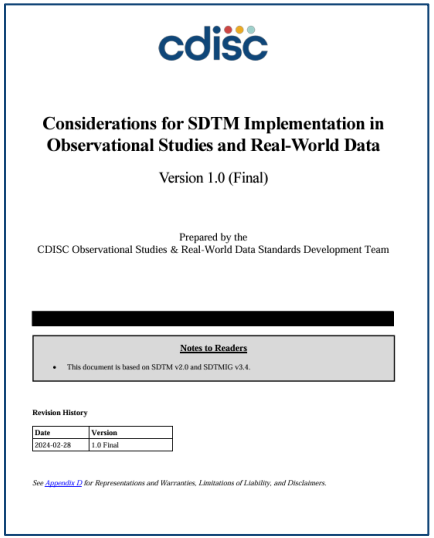
For questions regarding this draft document, contact (CDER) Dianne Paraoan, 301-796-2500, or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010.

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)
Oncology Center of Excellence (OCE)

February 2023

Real-World Data/Real-World Evidence (RWD/RWE)

- **Type of source for ECA** e.g., RCT, RWD
- **Access to Data**
- Design Consideration
- Analysis Considerations
- **No CDISC**


cdisc

Considerations for SDTM Implementation in Observational Studies and Real-World Data

Version 1.0 (Final)

Prepared by the
CDISC Observational Studies & Real-World Data Standards Development Team

Notes to Readers

- This document is based on SDTM v2.0 and SDTMIG v3.4.

Revision History

Date	Version
2024-02-28	1.0 Final

See Appendix D for Representations and Warranties, Limitations of Liability, and Disclaimers.

- Focus on **SDTM** only
- Cohort, Case Control and **ECA Studies**
- DM / TS consideration for ECA e.g., ARM
- How to handle conformance issues

- Focus on **ADaM** only
- Registry, case-control, etc. can be ECA, but **no specific discussion for ECA**
- Analysis visit Windowing
- Missing Imputation

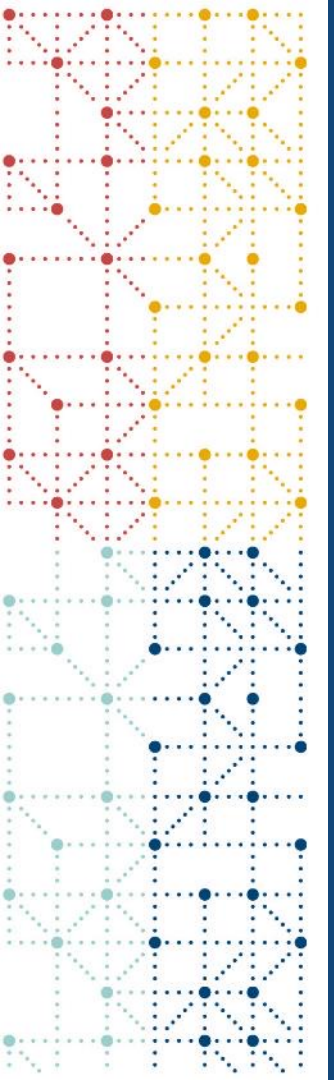
Current Guidance and Industry Initiatives

CDISC Initiative: RWD Lineage

An initiative to create a data exchange standard for lineage metadata that is supplied along with RWD-derived SDTM, which provides the data reliability required by FDA to use RWE as primary evidence

https://wiki.cdisc.org/display/RWDLIN/RWD+Lineage

Ref: A New CDISC Standard for Reliable Real World Data (RWD), PHUSE Real World Data Spring Event, April 2025



Case Study 1

Considering 2 RWE data with Randomised pivotal trial

Rationale

Filling Gaps in the Trial Data: External data helped provide information for time points or patient groups not fully covered in the main study.

Longer-Term Comparison: External patients were used to compare outcomes over a longer period, including after the main trial ended.

Understanding the Treatment Impact: This helps show how patients might do with and without treatment over time.

Making Fair Comparisons: We used matching methods to make sure the groups were similar at the start, so the results would be more reliable.

Data collected

Treated Pivotal Study (Randomized Controlled Trial)

- Phase III, randomized, double-blind, placebo-controlled trial with a long-term open-label extension.

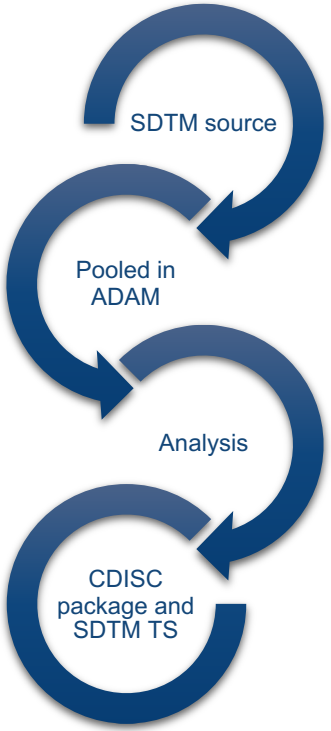
ECA Source A

- Multicenter longitudinal observational natural history cohort of patients with a similar population in region A

ECA Source B

- Multicenter longitudinal observational natural history cohort of patients with a comparable population in region B

- **Covariates Considered:**
  - Baseline functional score
  - Genetic subtype
  - Sex
  - Baseline age
  - Age at symptom
  - etc



Analytical & Technical Challenges

Data Cleaning and harmonization	General Programming	Platform & Runtime Limitations	Statistical Modeling Challenges	Regulatory Agility
•Unclean, missing and partial data. •Data cleaning possibility is limited. •Integrated TS domain across 3 studies (2 RWE study, 1 Pivotal) despite lacking information in iSAP. •Treatment vs. untreated subject data required harmonization.	•One of the RWE data had lacked SDTM compliance, demanded careful transformation for ADaM IG alignment. •Traceability challenges due to non-traditional SDTM structures. •Pinnacle 21 checks flagged missing DM/EX datasets – expected per FDA guidance for such RWE studies	•Some models took 4–5 hours per figure to run increasing turnaround time for even minor QC errors.	•Propensity score matching across studies with differing covariate definitions complicated model development. •Population models required repeated tuning new covariates had to be introduced after deep investigation by sponsor.	•Regulatory reviews required rapid responses, versioning, and documentation updates with minimal lead time.

Submission challenges

- One of the key RWE datasets was kept **blinded to maintain confidentiality**
- **Regulatory authorities showed strong interest** in the blinded data
- **Database holders were hesitant** to share data due to patient privacy concerns
- **Programs were modified** to run securely on a **third-party system**
- **Clear instructions and documentation** were provided for running the programs
- Maintained **close collaboration with the sponsor** throughout the process
- **Preparedness and flexibility** were essential to accommodate evolving data access requests

Case Study 2

The Comparative Effectiveness of Hydromethylthionine Mesylate (HMTM) Monotherapy in Subjects with Alzheimer’s Disease versus CPAD cohort based on propensity score matching

Rationale

- A small amount of an inactive compound was added to the placebo to mimic treatment effects (e.g., urine discoloration).
- This dose was expected to have no clinical activity based on earlier trial results.
- Since the trial did not include a fully inactive control group.
- This limits the ability to clearly assess the treatment's true effectiveness.
- Residual symptomatic effects may have confounded interpretation of the treatment's disease-modifying potential.

Data collected

Treated Pivotal Study

- Phase III, randomized, double-blind, placebo-controlled trial with a long-term open-label extension.
- SDTM and ADAM used

CPAD data

- The Critical Path for Alzheimer's Disease (CPAD) database contains a pool of patients coming from several Randomized Controlled Trials (RCTs).
- Data structure: SDTMs

- **Covariates Considered:**
  - Age,
  - Sex,
  - Smoking history,
  - Education,
  - Genotype,
  - Baseline MMSE,
  - etc

- Long process to get access to the data
- Pool of Metadata from various trial without proper metadata.
- It requires expert knowledge to fix the right parameters.
- ADAM like dataset generated.
- No CDISC package prepared
- Data published in scientific journal,
- will respond authorities along the phase3 with ECA results

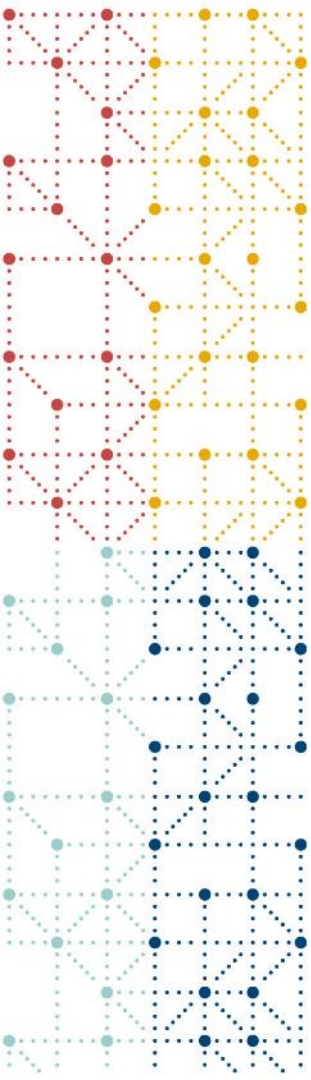
Data Harmonisation & Integration Challenges

Restricted Access

- 1 unique access to Pool Randomised TA third party database
- Data to be deleted after use
  - Prevents reproducibility of analyses

Data Challenges

- Difficulty to select **appropriate data** from RWE database:
- Many complex variables, requiring clinical expert interpretation
- This is a pool of several RCTs, Origin of data not always known with potential important information not collected as compared to more recent trials
- Difficulty to select **appropriate patients**:
- Applying inclusion/exclusion criteria of pivotal study to RWE database led to:
  - Complex algorithms
  - Manual medical review
- Difficulty to get **study assessments correspondence** between RWE database & pivotal study:
- Missing data for some of the outcomes
- Outcomes collected with different tools (e.g. different scoring versions)
- Outcomes collected at different timepoints leading to complex visit windowing



Learnings

Learnings and Future considerations

Mapping real-world data to SDTM/ADaM takes time and custom work

Design flexible CDISC structures to adapt diverse data sources

Align external and trial data to CDISC early in the process

Watch for differences in visit timing, assessments, and missing data

Handle non-standard IDs and missing dates systematically

Harmonize terminology (e.g., MedDRA, lab units) across sources

Learnings and Future considerations

- Creating ADaM datasets may require alternate derivations
- Ensure full traceability with clear documentation and define.xml
- Clearly document all assumptions, data limitations and transformation steps
- Validate pooled datasets carefully to catch small inconsistencies
- Plan for extra sensitivity analyses due to external data variability
- Be prepared for additional questions from regulatory

Thank You!

References

1	"**Data Standards for Drug and Biological Product Submissions Containing Real-World Data**" (FDA, 2023) https://www.fda.gov/media/153341/download
2	"**Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products** " (FDA, 2023) https://www.fda.gov/media/164960/download
3	"***CDISC Consideration for SDTM in Observational Studies and RWD***" (CDISC, 2024) https://www.cdisc.org/sites/default/files/2024-02/Considerations%20for%20SDTM%20Implementation%20in%20Observational%20Studies%20and%20Real-World%20Data%20v1.0.pdf
4	"***PHUSE Data Standards for Non-Interventional Studies***" (PHUSE, 2020) https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Data+Standards+for+Non-interventional+Studies.pdf

Abstract

The use of External Control Arms (ECA) in clinical trials is increasing, particularly for rare diseases where typical Randomized CTs may be difficult. Recent FDA guidance emphasizes both the potential and challenges of ECAs, emphasizing on data reliability, bias mitigation, adherence to CDISC SDTM and ADaM, statistical approaches such as propensity score matching, and regulatory communication. Additionally, CDISC and PHUSE have released guidance on integrating Real-World Data (RWD) into CDISC datasets.

In this presentation, we will summarize key insights from these documents regarding the use of RWD for ECAs and showcase two case studies on integrating ECA data into CDISC-compliant datasets

(1) constructing an ECA using natural history studies and past RCTs and
(2) leveraging publicly available RWD and RCT sources, such as the Critical-path Alzheimer's Disease database. We will discuss data integration, conformance challenges, and regulatory engagement, offering lessons for future rare disease studies.