Paper PP26

# Our Journey in Integrating External Control Arms (ECAs) and Real-World Data (RWD) for Rare Disease Trials

Cyril Sombrin, Cytel Inc., Geneva, Switzerland
Gautham Selvaraj, Cytel Inc., London, United Kingdom

## ABSTRACT

External Control Arms (ECAs) are increasingly applied in rare disease trials where Randomized Controlled Trials (RCTs) may be impractical. Recent guidance from the FDA, CDISC, and PHUSE highlights the importance of data reliability, bias reduction, adherence to standards (SDTM, ADaM), and the regulatory use of Real-World Data (RWD). This poster summarizes best practices and presents two case studies: constructing an ECA from natural history studies and legacy RCT data and leveraging publicly available RWD and RCT sources such as the Critical Path for Alzheimer's Disease (CPAD) database. We address common challenges in data integration, standardization, and statistical approaches such as propensity score matching, alongside practical strategies for regulatory communication and dataset conformance.

## INTRODUCTION

External Control Arms (ECA) offer you a powerful tool for enhancing clinical trial efficiency and regulatory acceptance. By utilizing real-world data (RWD) to construct comparator cohorts, ECAs provide valuable insights into treatment efficacy and safety. The use of ECAs is particularly valuable in therapeutic areas where placebo administration is unethical or where patient populations are extremely limited. ECAs can shorten development timelines, reduce costs, and support regulatory submissions when randomized control data are unavailable. This is often a challenge in rare disease trials, often facing recruitment limitations, ethical considerations, and logistical constraints. ECAs provide an alternative by leveraging historical clinical data, natural history studies, or Real-World Data (RWD) to replace or supplement a control group.

## REGULATORY CONTEXT AND CDISC STANDARDS AND PHUSE RECCOMENDATIONS

Recent FDA, CDISC guidance and PHUSE White Papers, emphasize the importance of data quality, transparent data lineage, bias mitigation, and adherence to clinical data standards when generating evidence from ECAs. This includes the FDA's "*Considerations for the Design and Conduct of Externally Controlled Trials*" [1] and "*Data Standards for Submissions Containing RWD*" [2], outline expectations for Clinical Trials using ECAs.

In [1] the FDA describes type of source for ECA e.g., RCT, RWD, etc. and way of accessing data as well as it provides some Statistical Analysis recommendation. In [2] the FDA provides more details and recommendations in providing datasets containing RWD and provides recommendations for creating define.xml and reviewer guide; the focus is only on SDTM. The guidance emphasizes the need for high-quality, well-documented RWD sources, as well as well-defined and consistent endpoints. Strong also emphasis on bias mitigation and CDISC alignment.

FDA is open-minded but cautious, ECAs are not a shortcut to approval

CDISC and PHUSE have also expanded standardization recommendations for RWD pipelines, including emerging work on RWD Lineage. With "*CDISC Consideration for SDTM in Observational Studies and RWD*" [3] and "*PHUSE Data Standards for Non-Interventional Studies*" [4]. While the CDISC Guidance focus is SDTM, the PHUSE white paper focuses more on ADaM. Some topics covered in the two guidance maybe relevant for ECA.

The CDISC guidance has some considerations for TS and DM, for example how to set the ARM variable, and provides some recommendations on how to handle conformance issues, given the fact implementing RWD might break some standard SDTM conformance rules for example. The PHUSE guidance, being its focus ADaM, provides recommendations and examples on applying in ADaM missing imputations and visit-windowing, as data coming from RWD might not have a precise visit schedule as in a traditional RCT.

Recently CDISC has launched a new Initiative, "RWD Lineage" ("A New CDISC Standard for Reliable Real World Data (RWD)", PHUSE Real World Data Spring Event, April 2025; https://phuse.s3.eu-central-1.amazonaws.com/Archive/2025/Webinar/Worldwide/Virtual/PRE_RWD06.pdf). The purpose of this initiative is to generate reliable RWD in SDTM for regulatory use, where additional information is needed to audit source data and quantify the information loss and performance of data transformations to calculate error bounds. RWD Lineage, will be a standardized and comprehensive representation of source data containing lineage for each source patient data element that specifies either, the location of the element in the output analysis dataset (Positive Lineage), or that the element was not used in the output analysis dataset (Negative Lineage). Furthermore, RWD Lineage will be represented in a standard metadata model that can be used along with SDTM to support the use of RWD with SDTM. This standardization will allow lineage to be combined across disparate data sources and will enable the unification of

downstream tools, workflows, and analytics for validation and audit activities. Use of RWD in ECA, is among one of the identified use cases planned in this initiative.

## CHALLENGES IN INCORPORATING ECAS

Despite their advantages, ECAs pose several methodological, operational, and data management challenges. These include:

- Data quality variation across sources
- Missing or inconsistent variables and visit schedules
- Lack of randomization and baseline imbalances
- CDISC conformance and traceability difficulties
- Programmatic complexity when working with blinded or externally hosted datasets

Furthermore, advanced statistical techniques such as propensity score matching (PSM), inverse probability weighting (IPW), and sensitivity analyses are often required to demonstrate robustness.

## SOURCE OF ECA DATA

An ECA is only as good as the data it relies on. Broadly, there are three main sources:

- Other clinical trials:
    - Prior trials of standard of care treatments can serve as external comparators.
    - Individual patient-level data (IPD) is preferred, but often only summary data is available.
    - These data are typically high quality but may not perfectly match the new study population.
- Published studies:
    - Systematic reviews and meta-analyses of literature can provide comparator data.
    - Useful when IPD is unavailable but limited by reporting standards and heterogeneity across studies.
- Real-world data (RWD):
    - Sources include electronic health records, registries, and insurance claims databases.
    - These capture routine clinical practice, reflecting the diversity of real patients.
    - However, RWD often suffers from missing data, variable quality, and lack of standardized endpoints.

Each source has strengths and weaknesses. Often, the best approach is to triangulate across multiple sources, ensuring that conclusions do not rest on a single dataset.

For more information, see Cytel blog “External Control Arms: A Powerful Tool for Oncology and Rare Disease Research” (https://cytel.com/perspectives/external-control-arms-a-powerful-tool-for-oncology-and-rare-disease-research/).

## CYTEL CASE STUDIES

We want to report two cases where Cytel Statistical Programming team was involved in. Each use case has its own peculiarity and challenges.

### CASE STUDY 1: ECA FROM RWD FOR PIVOTAL PHASE III RCT

This was a Phase III, randomized, double-blind, placebo-controlled trial with a long-term open-label extension.
The ECA made use of two RWD natural history cohorts. The ECA filled information gaps in the pivotal trial, enabling long-term comparisons and better understanding of treatment effects. The challenges encountered included:

- harmonization of functional scores
- matching procedures
- ensuring dataset traceability.

Technical challenges included:

- inconsistent visit schedules
- missing baseline variables
- variations in genetic stratification across datasets

Analytical challenges required robust matching methodologies, multiple sensitivity analyses, and validation of assumptions around exchangeability.
One of the RWD sources had lacked SDTM compliance, demanding careful transformation for ADaM IG alignment.
Traceability was also challenging due to non-traditional SDTM structures.
Pinnacle 21 checks flagged missing DM/EX datasets, expected per FDA guidance for such RWE studies; all conformance issues were reported and rationale provided in the review guide that was part of the submission package.
Among the issues, one of the key RWD datasets was kept blinded to maintain confidentiality. The RWD holders were hesitant to share data due to patient privacy concerns. However, the Regulatory Authorities showed strong interest in this blinded data. We then decided to provide SAS programs running our analysis on the provider platform, by providing

clear instructions and documentation for running the programs. The provider was then responsible for sharing results from this additional directly with the Regulatory Authorities.

#### CASE STUDY 2: USING CPAD FOR COMPARTIVE EFFECTIVENESS

Case Study 2 was Phase III pivotal studies, comparing Hydromethylthionine Mesylate (HMTM) with patients from the Critical Path for Alzheimer's Disease (CPAD) database (https://c-path.org/program/critical-path-for-alzheimers-disease/).

The CPAD dataset consisted of pooled RCT data in CDISC SDTM format from multiple studies, resulting in variability in SDTM implementation, coding dictionaries, visit naming conventions, and metadata completeness. Harmonization required a multi-step approach including mapping, crosswalk creation, and rigorous QC.

One major challenge in accessing this specific data source was CPAD requiring you to delete data after use, possibly preventing reproducibility of analyses at the Regulatory Authorities.

Furthermore, from the technical side we had the following issues:

- Difficulty in selecting appropriate data from RWE database:
- Many complex variables, requiring clinical expert interpretation
- This is a pool of several RCTs, and therefore the origin of the data was not always known with potential important information not collected as compared to more recent trials
- Difficulties in selecting appropriate patients
- Applying inclusion/exclusion criteria of pivotal study to RWD database led to Complex algorithms and a deep manual medical review
- Difficulty in getting study assessments correspondence between RWD 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 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 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
- 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 regulators.

## CONCLUSION

ECAs offer a scientifically credible and operationally efficient approach to supplement evidence generation when traditional clinical trial designs are not feasible. The case studies presented demonstrate that with careful data selection, robust analytical methods, and adherence to CDISC and regulatory expectations, ECAs can meaningfully support regulatory submissions and real-world decision-making.

## REFERENCES

1. "Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products" (FDA, 2023) https://www.fda.gov/media/164960/download
2. "Data Standards for Drug and Biological Product Submissions Containing Real-World Data"(FDA, 2023) https://www.fda.gov/media/153341/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

## ACKNOWLEDGMENTS

We would like to thank Angelo Tinazzi for his support in writing and reviewing the poster and drafting this paper.

## RECOMMENDED READING

If you want to know more about ECA, see Cytel blogs here:

- External Control Arms: A Powerful Tool for Oncology and Rare Disease Research (https://cytel.com/perspectives/external-control-arms-a-powerful-tool-for-oncology-and-rare-disease-research/)
- The Role of External Data in Oncology Drug Development (https://cytel.com/perspectives/the-role-of-external-data-in-oncology-drug-development/)

## CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Author Name: Cyril Sombrin
Company: Cytel Inc.
Address: Geneva
Email: Cyril.Sombrin@cytel.com
Website: www.cytel.com