Generating CDISC-Compliant Data from LEO CReWE for Clinical Trial Development

Michael J. Kane, Dai Chihara, Amy Ayers, and Christopher Flowers
MD Anderson Cancer Center
*Corresponding author: mjkane@mdanderson.org

Background and Introduction

- Regulatory agencies (**FDA**, **EMA**) require **CDISC**-formatted datasets for submissions.
- **SDTM** organizes source data; **ADaM** prepares analysis-ready datasets with full traceability.
- Oncology trials are complex — multiple endpoints, longitudinal follow-up, evolving therapies.
- Academic real-world datasets (RWD) are rarely CDISC-compliant, limiting their use in:
    - Trial planning and monitoring
    - Regulatory submissions
- **LEO CReWE**: Multi-institutional registry of lymphoma characteristics, treatments, and outcomes; has supported submissions in relapsed/refractory DLBC and FL.

Objectives

1. Create CDISC-Compliant Datasets
- Generalized **SDTM** for DLBC and FL cohorts.

2. Develop Protocol-Specific ADaM
- Case studies: R-GemOx in ≥2L DLBC; mosunetuzumab in R/R FL.

3. Ensure Traceability & Validation
- Clear mapping from source data to submission datasets.

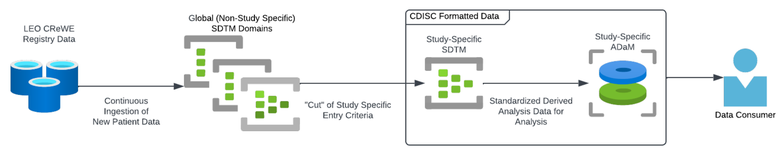
4. Enable Scalability
- Workflow adaptable to multiple oncology protocols.

Methods

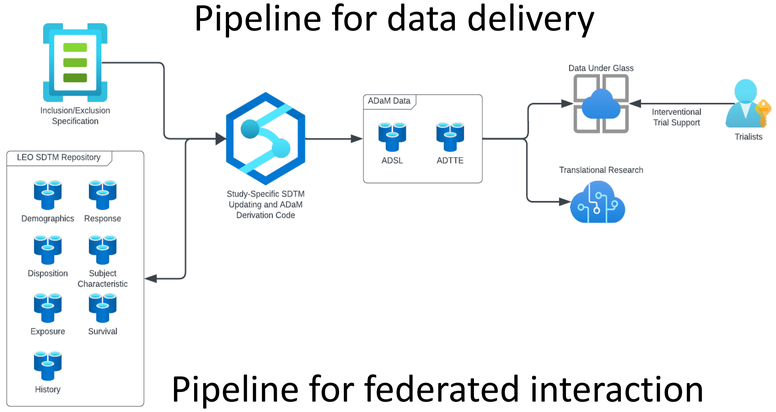
Workflow Overview

1. Extract LEO CReWE source data (clinical, treatment, outcome variables).
2. Transform into a **generalized SDTM** structure using customized open-source tools.
3. Map protocol-specific endpoints to **ADaM** datasets.
4. Validate for compliance, traceability, and reproducibility.

Key Features
- Automated mapping templates to reduce manual coding.
- Traceability links from raw data to submission variables.
- Modular design enabling reuse for multiple oncology studies.
- Standardization provides opportunities for federated interaction



Pipeline for data delivery

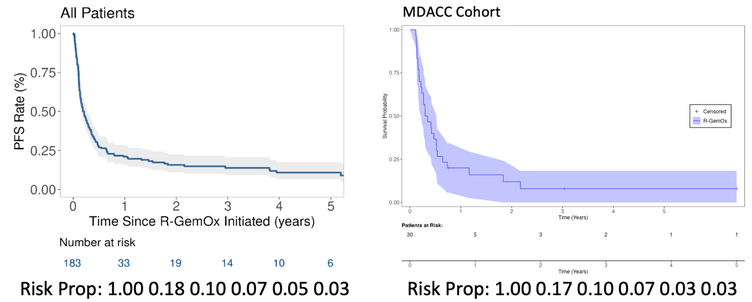


Pipeline for federated interaction

Case Study: GemOx >= 2L DLBC[1]

Cohort
- Patients with **diffuse large B-cell lymphoma** (DLBC)
- Received **R-GemOx** in **second-line or later** settings
- Identified from **MD Anderson LEO CReWE data** (within multi-institutional, longitudinal RWD)

Workflow
- **Extracted** relevant source data.
- **Mapped** to **generalized SDTM domains**
- **Generated** protocol-specific **ADaM** datasets (ADSL, ADTTE)



PFS GemOx[1] Entire Cohort and MDACC Cohort

Conclusion and Clinical Impact

Workflow converted LEO CReWE data into **CDISC-compliant SDTM and ADaM**.

Scalable for multiple oncology protocols.

Impact
- **Faster** trial planning and analyses.
- **Bridges** academic RWD with regulatory needs.
- **Enables** broader use of real-world evidence in oncology development.

* Yamshon, Samuel, et al. "Outcomes of Relapsed or Refractory Diffuse Large B-Cell Lymphoma Treated With R-GemOx: A Multicenter Cohort Study." *American journal of hematology* 100.4 (2025): 606-615.