PharmaSUG 2026 - Paper RW-248

Real-World Data and CDISC – An Evolving Journey

Shuo (Serena) Cao, Ph.D., Ephicacy Consulting Group;

Venkat Rajagopal, Ph.D, Ephicacy Consulting Group

ABSTRACT

Real-world data (RWD) and real-world evidence (RWE) have become essential components in modern clinical research, regulatory decision-making, and healthcare delivery. RWD, originating from electronic health records (EHRs), insurance claims, patient registries, and digital health tools, captures patient experiences in routine clinical practice. Unlike traditional randomized clinical trials (RCTs), which operate under controlled protocols and narrowly defined populations, RWE provides insights into treatment effectiveness, safety, and utilization across heterogeneous populations in real-world settings.

The integration of RWD into regulatory and research workflows presents several challenges, including structural heterogeneity, semantic differences, and varying data quality. To address these challenges, interoperability frameworks, controlled terminologies, and standardized data models are critical. The Clinical Data Interchange Standards Consortium (CDISC) has traditionally provided clinical research standards; however, the growing use of RWE has prompted a shift toward co-evolutionary frameworks, where RWE informs CDISC standards adaptation. This paper explores this evolution, emphasizing interoperability as the enabler, implications for regulatory-grade RWE, and future directions toward standards-ready ecosystems.

INTRODUCTION

Over the last two decades, the healthcare industry has undergone a digital transformation that has revolutionized data capture, storage, and analysis. Electronic health records (EHRs), claims databases, disease registries, and digital health technologies now provide longitudinal patient data across diverse populations. These datasets reflect real clinical care, capturing disease progression, treatment patterns, and patient outcomes in contexts not always observable in controlled clinical trials.

Randomized controlled trials remain the cornerstone for establishing drug efficacy and safety. They provide rigorous internal validity through standardized protocols, randomization, and strict inclusion/exclusion criteria. However, these trials often exclude patients with multiple comorbidities, the elderly, or those receiving combination therapies, limiting their generalizability. Moreover, the controlled setting rarely reflects the variability of real-world clinical practice.

RWE complements clinical trials by providing insights into broader populations, longer follow-up periods, and naturalistic treatment pathways. Regulators, including the FDA and EMA, increasingly acknowledge the value of RWE for label expansion, post-marketing surveillance, and comparative effectiveness studies. Nonetheless, leveraging RWD for research requires overcoming substantial challenges related to data standardization, interoperability, and quality assurance.

CHALLENGES OF REAL-WORLD DATA STANDARDIZATION

The primary challenge of real-world data is its heterogeneity. Healthcare systems use diverse EHR platforms, coding systems, and workflows. Within a single institution, data may be fragmented across laboratories, pharmacy systems, radiology platforms, and administrative databases. This structural variation complicates aggregation and analysis.

Semantic heterogeneity further amplifies the complexity. Clinical concepts may be coded in ICD, SNOMED CT, LOINC, or RxNorm. Mapping these disparate terminologies into a unified framework requires domain expertise and robust governance. Inconsistent units, free-text entries, and missing values further undermine data reliability.

Data quality is a persistent issue in RWD. Unlike RCTs, where data collection is protocol-driven, RWD reflects operational priorities, leading to missing, incorrect, or inconsistent data. For example, lab measurements may not follow uniform timing or units, medications may be recorded differently across sites, and clinical notes may contain ambiguous narrative information. Ensuring reproducibility requires standardized transformation and rigorous quality assessment.

Addressing these challenges necessitates semantically interoperable frameworks, controlled vocabularies, and metadata-driven governance processes. By aligning disparate datasets through harmonization, organizations can produce reliable, regulatory-grade real-world evidence.

FROM TRANSFORMATION TO CO-EVOLUTION: RWE AND CDISC

CDISC has traditionally provided one-way mapping standards for clinical trial data. SDTM (Study Data Tabulation Model) organizes raw trial data into standardized domains, while ADaM (Analysis Data Model) facilitates consistent statistical analysis. Historically, RWD was transformed into these frameworks through extract-transform-load pipelines, primarily to satisfy regulatory submission requirements.

While transformation enables compliance, it often discards clinical context. Real-world interactions, such as complex treatment pathways or comorbidity management, may not conform to rigid CDISC structures. As RWE becomes central to regulatory decisions, a co-evolutionary approach emerges: RWE informs and refines CDISC standards. Rather than a static endpoint, CDISC evolves alongside RWD methodologies, incorporating new domains, metadata, and analytic conventions that capture real-world complexity.

Co-evolution ensures that standards remain relevant for emerging evidence types. For example, adaptive data models can now accommodate wearables, remote monitoring, and patient-reported outcomes. By iteratively integrating RWE insights, CDISC evolves into a living framework capable of supporting both traditional trials and real-world studies.

INTEROPERABILITY AS THE ENABLER

Interoperability is the cornerstone of effective RWD utilization. HL7 FHIR (Fast Healthcare Interoperability Resources) is a widely adopted standard that defines modular, API-driven healthcare resources, enabling seamless data exchange across EHRs, registries, and research platforms. FHIR resources represent patients, encounters, procedures, observations, and medications with standardized formats.

Controlled terminologies ensure semantic consistency. SNOMED CT standardizes diagnoses, LOINC codifies laboratory results, and RxNorm harmonizes medications. Integrating these standards with CDISC models facilitates bidirectional mapping while preserving clinical context and regulatory traceability.

Advanced interoperability allows automated extraction, harmonization, and transformation pipelines that maintain longitudinal patient information. By preserving source data fidelity, organizations can generate high-quality RWE suitable for submission and cross-study analysis.

IMPLICATIONS FOR REGULATORY-GRADE RWE

Standardized, interoperable RWD improves reproducibility, credibility, and regulatory confidence. By aligning real-world datasets with CDISC and interoperability frameworks, researchers can trace analytical results to their source, satisfy audit requirements, and demonstrate transparent methodology.

Moreover, harmonization enables data reuse. The same dataset can support pharmacovigilance, comparative effectiveness, and health economics analyses, maximizing resource efficiency. Standardized RWE accelerates study timelines and facilitates multi-center collaborations, enhancing the translational value of research.

Ultimately, regulatory-grade RWE bridges research, care delivery, and policy, providing actionable insights that complement traditional clinical trials while maintaining scientific rigor.

FUTURE DIRECTIONS AND CONCLUSION

The future of RWE generation is a fully integrated ecosystem, combining interoperable healthcare systems, standardized research frameworks, and intelligent automation. Machine learning and metadata-driven pipelines will further streamline data transformation, quality assessment, and analytic readiness.

Embedding standards at data ingestion, rather than applying them retrospectively, reduces error propagation and supports scalable RWE studies. Adaptive frameworks will continue to evolve with novel data sources, including wearables, genomics, and patient-reported outcomes.

The convergence of RWD, RWE, and CDISC standards represents a methodological inflection point. By embracing interoperability, semantic harmonization, and co-evolutionary standards, the life sciences industry can generate transparent, reproducible, and patient-centered evidence that informs regulatory and clinical decisions.

REFERENCES

1. Corrigan-Curay, J., Sacks, L., & Woodcock, J. (2018). Real-world evidence and real-world data for evaluating drug safety and effectiveness. *Clinical Pharmacology & Therapeutics.*
2. Sherman, R. E., Anderson, S. A., Dal Pan, G. J., et al. (2016). Real-world evidence — What is it and what can it tell us? *New England Journal of Medicine*.
3. Kahn, M. G., Brown, J. S., Chun, A. T., et al. (2016). Transparent reporting of data quality in distributed data networks. *eGEMs*.
4. Kostkova, P., Brewer, H., de Lusignan, S., et al. (2016). Who owns the data? Open data for healthcare. *Journal of Medical Internet Research*.

5. Kahn, M. G., Callahan, T. J., Barnard, J., et al. (2021). A harmonized data quality assessment terminology and framework for the secondary use of electronic health record data. *eGEMs.*