White Paper

Demystifying data requirements for RWE in regulatory submissions

Shivani Aggarwal, PhD, MS, CEO, Landmark Science, Inc.
Barath Sukumar, MS, Senior Statistical Programmer, Landmark Science, Inc.
David Goldfarb, PhD, MPH, Senior Manager, Epidemiology, Landmark Science, Inc.

https://landmarkscience.com

Table of Contents

1. Overview

The regulatory framework for real-world evidence (RWE) has evolved substantially over the past decade, particularly following the 21st Century Cures Act of 2016. The acceptance of RWE by regulators has led to its increased industry adoption and the development of additional guidelines and consensus on its application. Numerous guidance documents for industry have been set forth by the United States (US) Food and Drug Administration (FDA) on the use of real-world data (RWD) and RWE, each of which detail various considerations for RWE in regulatory decision-making, from the selection of data sources such as electronic health records (EHRs), claims, and registries, to considerations for conducting non-interventional studies and externally controlled trials, to data standards and submission requirements – a crucial yet often overlooked aspect. In fact, recent FDA guidance emphasizes that RWD/RWE used in regulatory submissions must be submitted electronically and adhere to the approved data standards for such submissions. Not adhering to these guidelines on data standards may have implications on the regulatory submission and review process, such as delays or complications. Therefore, aligning RWD/RWE with the FDA's requirements is crucial to facilitate a streamlined regulatory review for submissions containing RWD/RWE.

Study data for clinical studies

Submission of study data has been a regulatory requirement and the industry standard for over two decades for Sponsors submitting clinical data in Investigational New Drug (INDs) applications, New Drug Applications (NDAs), Abbreviated NDAs (ANDAs), and Biologic License Applications (BLAs). Submission of study data, such as raw and analytic datasets, are required by regulators to demonstrate transparency and traceability in studies. These data may also be used to replicate analyses conducted by the Sponsor and enable regulators to rigorously evaluate the data through sensitivity analyses, redefining index dates, or alternate methodology(-ies).

RWD conformance is a new domain

Conformance and standardization of RWD into currently accepted data standards is still a relatively new domain for many pharmaceutical and biotechnological companies and poses a variety of challenges. The complexity of integrating RWD/RWE into the regulatory framework requires a deep understanding of relevant guidance documents, nuances of data standards, and potential obstacles that companies may face. Sponsors must navigate issues such as ensuring data quality and completeness and overcoming integrating diverse RWD into a unified format.

In this first part in a series of white papers on demystifying data standards for RWD/RWE in regulatory submissions, we explore key points from relevant FDA guidance documents to illustrate **best practices and actionable strategies** that Sponsors can apply when preparing their regulatory submissions.

Key Objectives

1) Summarize relevant FDA guidance documents
2) Describe currently accepted data standards
3) Summarize key elements of the data submission
4) Discuss key challenges that companies may encounter for RWD

2. Regulatory Background on Data Standards

Study data should be submitted electronically

Over the past two decades, the FDA has put forward numerous guidance documents related to clinical trials, including detailed guidance on data submission requirements to the FDA. In the more recent but seminal guidance document for industry *Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act* (original release: Feb 2014, final release: Dec 2014), the FDA specifies that data must be submitted in an electronic format.[1] The subsequent guidance document *Providing Regulatory Submissions In Electronic Format — Standardized Study Data* (original release: Feb 2014, latest release: Jun 2021) requires that data submissions be in a format the FDA can process, review, and archive.[2,3] Furthermore, the guidance states that data not submitted in this format will not be filed or received, unless the Sponsor has sought - and the FDA has granted - a waiver.

Data should comply with Data Catalog

This guidance document recommends that electronically submitted data comply with formats outlined in the FDA Data Standards Catalog (v10.3, latest release: Apr 2024), which contains data standards currently supported by the FDA for clinical and non-clinical data.[4] The Data Standards Catalog is a live repository of current and previously supported study data standards (e.g., ADaM, SDTM), exchange format standards, and controlled terminology standards (e.g., MedDRA for adverse events, LOINC for laboratory test names, and WHO Drug Dictionary for medications), and is periodically reviewed and updated by the FDA. In this white paper, we focus on study data and exchange format standards for RWD.

Regulatory Requirements for Data in Submissions

- Applies to data submitted for INDs, (s)NDAs, ANDAs, and (s)BLAs
- Data should be submitted in electronic format
- Data should comply with formats supported in the FDA Data Standards Catalog

(s) denotes supplemental

3. CDISC Data Standards Accepted by the FDA

About CDISC

While various data standards may be utilized in clinical research, data standards supported by the FDA are primarily those established by the Clinical Data Interchange Standards Consortium (CDISC). CDISC is an international consortium dedicated to establishing standards for clinical and nonclinical data. CDISC standards enable the mapping of data in a uniform, one-step-away programming, ensuring transparency and traceability to regulators such as the FDA.[5]

CDISC standards are required

CDISC (www.cdisc.org) standards are required for regulatory submissions to the FDA (US) and PMDA (Japan), and specify study data standards across all phases of a clinical study, from data collection in the eCRF to mapped source data, and to mapped analytic data. We summarize the CDISC data standards used throughout a clinical study below: Clinical Data Acquisition Standards Harmonization (CDASH), Study Data Tabulation Model (SDTM), and Analysis Data Model (ADaM).[6,7,8]

Table 1. Approved Data Standards Used in Clinical Studies

	Raw Data	Source Data	Analysis Data
CDISC Data Standard	CDASH[6]	SDTM[7]	ADaM[8]
Purpose	Used for collecting clinical raw data such as from electronic case report forms (eCRFs)	Used for creating source datasets from the raw data. SDTM datasets contain standardized variables.	Used for creating analysis datasets from source SDTM datasets. ADaM datasets contain derived variables for analysis and reporting of study results.
Mapping	Blank eCRF are annotated to create the aCRF (annotated CRF), which contains mapped CDASH variable names.	Raw data from the eCRF or CDASH is mapped to SDTM using the SDTM Implementation Guide (SDTMIG).[9]	Source data, such as SDTM datasets, are extracted and transformed to create analytic datasets using the ADaM Implementation Guide (ADaMIG).[10]
Example Datasets[†]	○ CRF forms	○ DM: Demographics ○ AE: Adverse Events	○ ADSL: subject-level analysis dataset ○ ADAE: adverse events analysis dataset

ADaM = Analysis Data Model, CDASH = Clinical Data Acquisition Standards Harmonization, SDTM = Study Data Tabulation Model. [†]For an exhaustive list, please refer to the respective implementation guides.

CDISC SDTM and ADaM are accepted FDA data standards

Of the standards for clinical data discussed above, SDTM and ADaM are specified in the Data Catalog and required for submission (**Table 2**).[4] Both the SDTM and ADaM data are generally submitted to the regulatory agency in SAS Transport File Format (XPT) along with the Define document containing the metadata in XML (Extensible Mark-up Language) format.[11]

It is recommended that Sponsors submit data used for INDs, (s)NDAs, ANDAs, and (s)BLAs in these formats. While the FDA does not prohibit the use of other data standards, alignment should be gained between the Sponsor and FDA – should the Sponsor wish to submit other data standards to the FDA – *prior* to the submission of the data in the alternate data standard.

Data Standards for Clinical Submissions

- SDTM for tabulation datasets
- ADaM for analysis datasets

Table 2. Summary of Data Standards Accepted by the FDA for Clinical Data

Use	Data Standard	Exchange	Standards Development Organization	Suggested for RWE Submissions
Clinical study datasets	SDTM	XPT	CDISC	✓
Clinical study datasets	ADaM	XPT	CDISC	✓
Study data definition	Define	XML	CDISC	✓

ADaM = Analysis Data Model, SDTM = Study Data Tabulation Model, XPT = SAS Transport File, a file format used to support data transfers, XML = Extensible Markup Language.

4. FDA Guidelines for RWD/RWE Data Submissions

Although the regulatory landscape for RWD/RWE is less evolved than that for clinical trials, the FDA has, in recent years, released numerous draft and final guidance documents detailing considerations for the use of RWD/RWE in regulatory decision-making. In particular, guidance documents have been issued that detail the FDA's thinking on the use of RWE for regulatory submissions. In the FDA's final guidance for industry *Data Standards for Drug and Biological Product Submissions Containing Real-World Data* (released Dec 2023), the FDA states that RWE submitted in filings such as NDAs or BLAs must be submitted electronically and comply with data formats specified by the FDA.[12]

Indeed, this RWE guidance references the above-mentioned guidance document *Providing Regulatory Submissions in Electronic Format—Standardized Study Data* (Study Data Guidance) (detailed in **Section 2**), thus suggesting that RWE submitted to regulators should be mapped to CDISC standards as specified in **Section 3**. The table below enumerates helpful references, including relevant FDA guidance documents for industry and CDISC guidance on data standards.

Table 3. Summary of Relevant Guidance Documents for RWE Submissions

Relevant Document	Latest Release Date
FDA Guidance for Industry: Data Standards for Drug and Biological Product Submissions Containing Real-World Data[12]	December 2023
CDISC: Considerations for SDTM Implementation in Observational Studies and Real-World Data[13]	February 2024
Study Data Technical Conformance Guide[14]	March 2024
FDA Guidance for Industry: Providing Regulatory Submissions in Electronic Format — Standardized Study Data[2]	June 2021
FDA Data Standards Catalog[4]	April 2024

5. Elements of Data Submissions to FDA

Key elements of data submission: standardized data, data dictionaries, and Reviewer's Guides

Here we outline the key elements typically included in a standardized data submission.[12,14,15] These elements include standardized datasets, detailed data dictionaries, and Reviewer's Guides, all of which play a pivotal role in supporting the FDA's evaluation process. By thoroughly preparing and submitting these components, Sponsors can facilitate a smoother review process, providing the FDA with the necessary context, clarity, and details to assess data such as RWD/RWE in the submission effectively.

Table 4. Elements of Data Submissions

Data Standard	Define Files	Data dictionary / Dataset specifications	Reviewer's Guide
Supported data standard	*Specifications that describe metadata for datasets, including variables, possible values, and controlled terminologies and codes*	*Data dictionary / dataset specifications that describe programming definitions for derived variables (i.e. logic).*	*Comprehensive documentation on study/analysis tables, conformance findings, and other helpful details for FDA review.*
SDTM standardized source datasets **ADaM** standardized analytic datasets	A **define.xml** file that contains metadata for each SDTM/ADaM dataset	**Data dictionary**: - Variables names and labels - Formats - Variable definitions - Controlled terminology **Dataset specifications:** - Dataset overview - Dataset structure - Variable derivations - Flags	**Reviewer's Guide (SDRG/ADRG)**: - Mapping information - Study objectives - Summary of datasets - Hardcodes implemented - Special data considerations - Summary of conformance findings

* ADRG = Analysis Data Reviewer's Guide, SDRG = Study Data Reviewer's Guide.

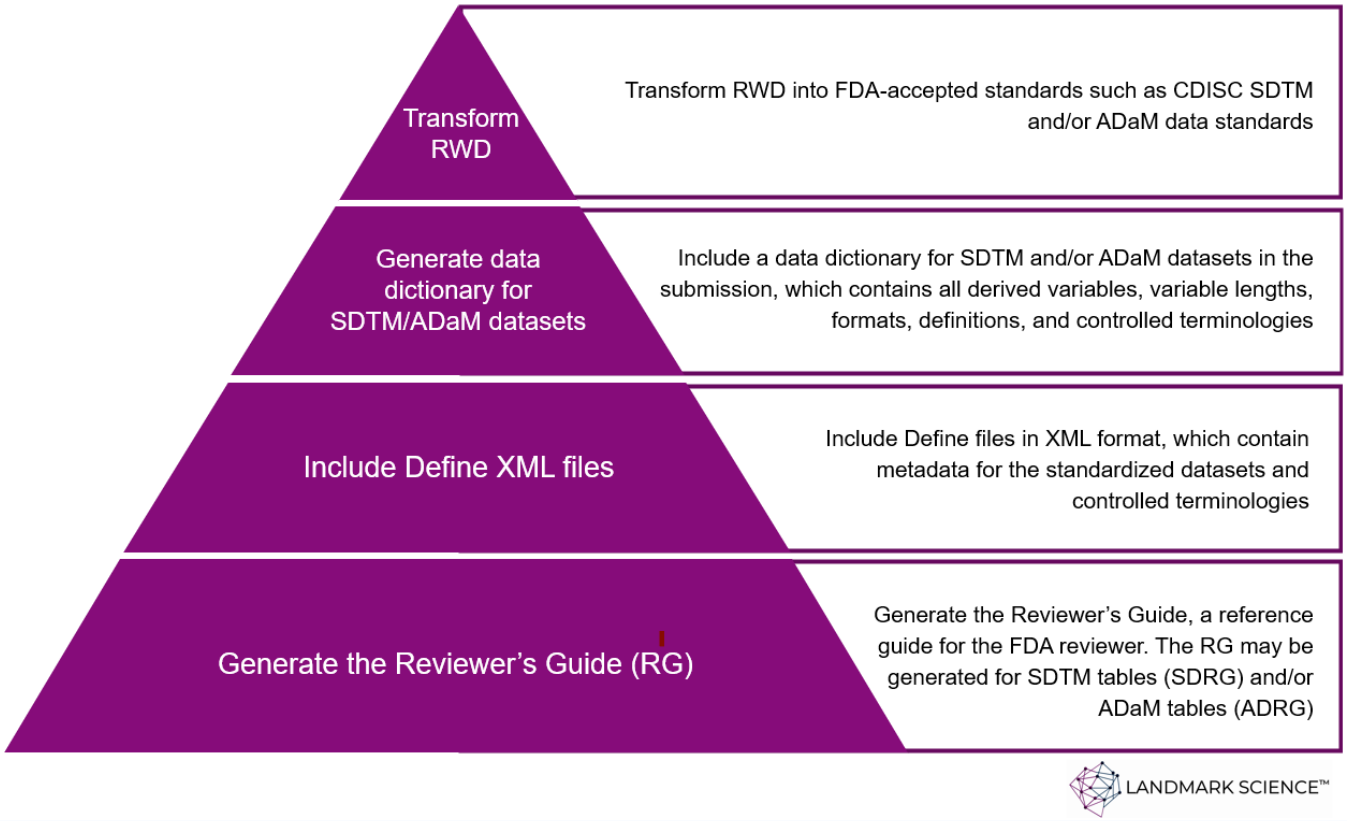
Reviewer's Guide: A Brief Primer

- A *single summary document* that serves as a reference guide for the FDA reviewer, the RG should describe:
  - available datasets (e.g., tabulation or analytic), special considerations or directions, conformance issues identified, hardcodes, and any other items that may facilitate the reviewer's use of the submitted data
- The Reviewer's Guide also assists the reviewer in understanding the relationships between the study report and the data
- Sponsors may use templates to complete the SDRG and ADRG

6. Core Aspects of RWD Submissions

When submitting RWD/RWE to the FDA, it is recommended that the data submission conforms to the required standards and that appropriate documentation is in place to ensure adherence to FDA guidelines. Comprehensive documentation not only aids in compliance but also enhances the transparency and interpretability of the submitted data. While components suggested for RWD submissions may vary on a case-by-case basis, in general the FDA recommends the following when submitting RWE for regulatory decision-making:

Figure 1. Elements of RWD Submission

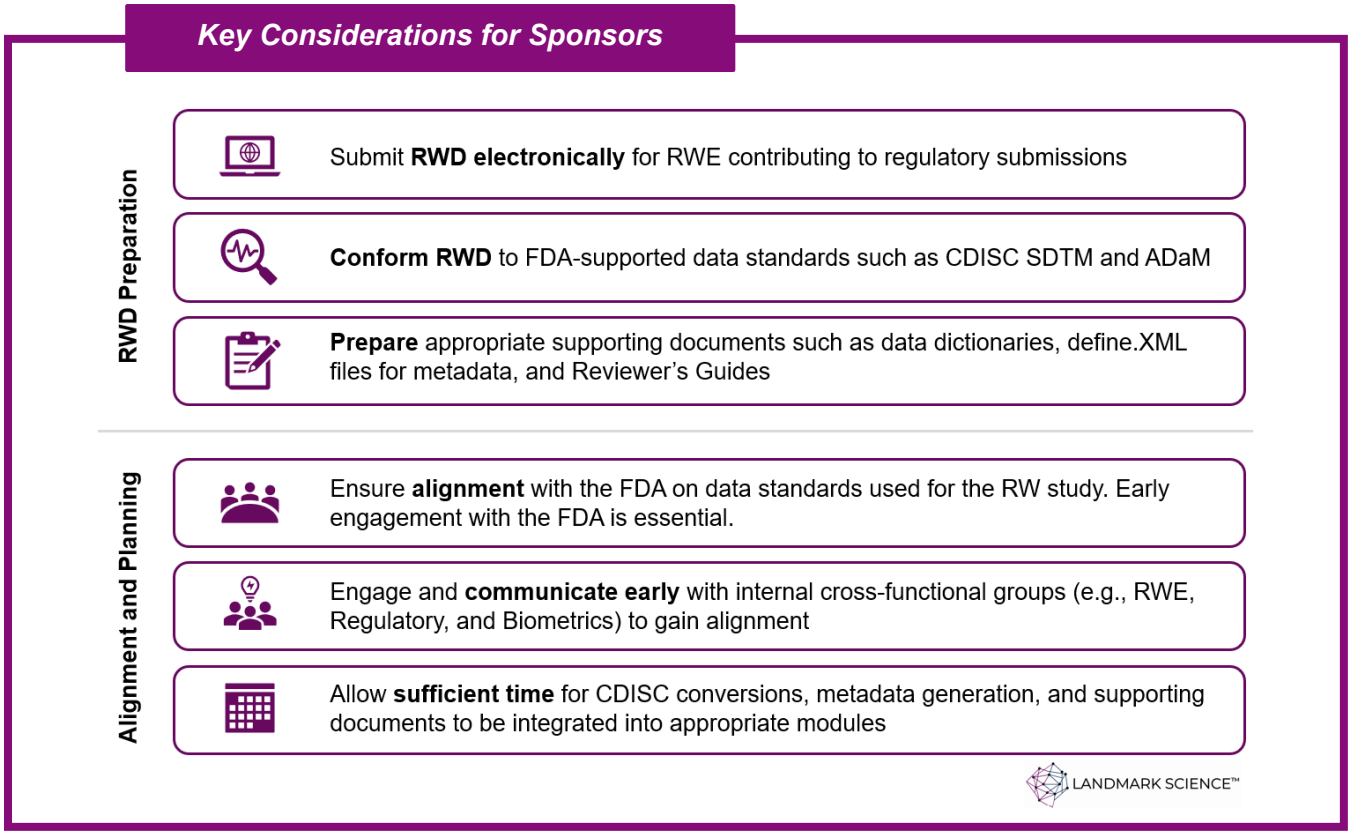


Challenges in RWD Conversion

Sponsors may encounter challenges when mapping RWD to data standards such as SDTM and ADaM. Conversions to CDISC standards can be challenging because of inherently different data collection processes, differing data standards used for RWD (e.g., Observational Medical Outcomes Partnership [OMOP]), and other challenges (e.g., missing data, lack of treatment assignment) used in observational research compared to clinical trials. These differences may be at the source level or at the analysis dataset level.

7. Summary

While data submissions in compliant formats have been the standard for clinical data for over two decades, such standards for RWD/RWE used in regulatory submissions are relatively new. Sponsors will need to navigate the complexities of recent RWD/RWE guidance documents and submission requirements when considering submitting RWD/RWE in regulatory submissions.



Stay Tuned for Part II of White Paper

In Part II of our white paper, *Demystifying Data Requirements for RWE in Regulatory Submissions*, we delve into the challenges that Sponsors may encounter during RWD conformance, offering potential solutions to address them.

8. References

1. FDA. *Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act.* U.S. Food and Drug Administration. Accessed August 2024.
2. FDA. *Providing Regulatory Submissions in Electronic Format — Standardized Study Data.* U.S. Food and Drug Administration. Accessed August 2024.
3. Federal Register. *Providing Regulatory Submissions in Electronic Format-Standardized Study Data; Guidance for Industry; Availability*. Federal Register: A Daily Journal of the United States Government. Accessed August 2024.
4. FDA. *Data Standards Catalog.* U.S. Food and Drug Administration. Version 10.3, Accessed April 2024.
5. CDISC – https://www.cdisc.org. Clinical Data Interchange Standards Consortium. Accessed August 2024.
6. CDISC CDASH – https://www.cdisc.org/standards/foundational/cdash. Clinical Data Interchange Standards Consortium. Accessed August 2024.
7. CDISC SDTM – https://www.cdisc.org/standards/foundational/sdtm. Clinical Data Interchange Standards Consortium. Accessed August 2024.
8. CDISC ADaM – https://www.cdisc.org/standards/foundational/adam. Clinical Data Interchange Standards Consortium. Accessed August 2024.
9. CDISC. *SDTM Implementation Guide.* Clinical Data Interchange Standards Consortium. Accessed August 2024.
10. CDISC. *ADaM Implementation Guide v1.3.* Clinical Data Interchange Standards Consortium. Accessed August 2024.
11. CDISC. *Define-XML.* Clinical Data Interchange Standards Consortium. Accessed August 2024.
12. FDA. *Data Standards for Drug and Biological Product Submissions Containing Real-World Data.* U.S. Food and Drug Administration. Accessed August 2024.
13. CDISC: *Considerations for SDTM Implementation in Observational Studies and Real-World Data.* Clinical Data Interchange Standards Consortium. Accessed August 2024.
14. FDA. *Study Data Technical Conformance Guide — Technical Specifications Document.* U.S. Food and Drug Administration. Accessed August 2024.
15. FDA. *Study Data Standards Resources.* U.S. Food and Drug Administration. Accessed August 2024.

9. About Landmark Science

Landmark Science is an epidemiology consulting company specializing in advanced statistical methods and analytics for RWE generation. Landmark Science conducts regulatory-grade studies and converts RWD to FDA-compliant data standards (e.g., CDISC) for regulatory submissions. For more information, please contact info@landmarkscience.com.