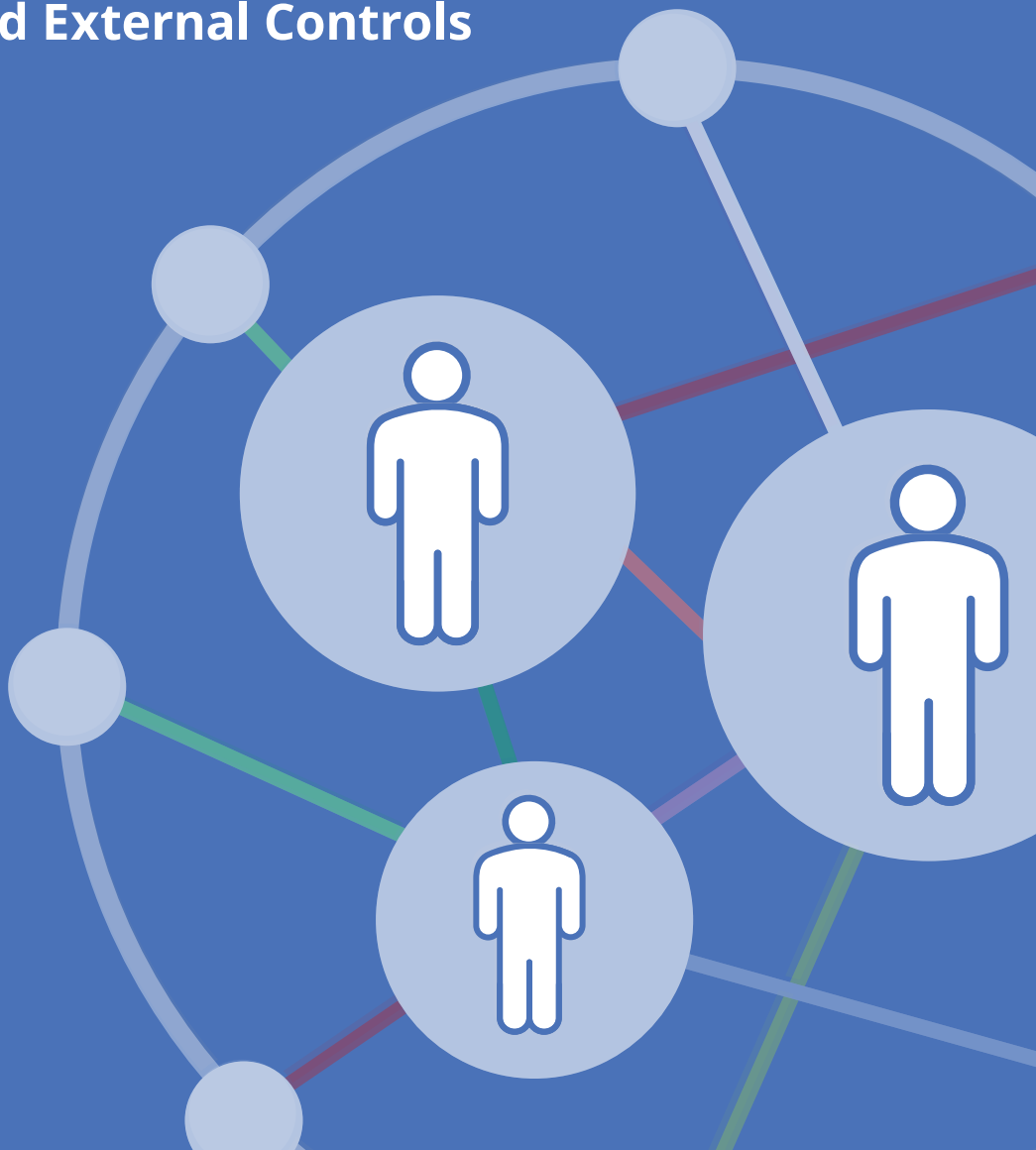
CRDSA

Clinical Research Data Sharing Alliance

Data Reuse in Regulatory Submissions: The Role of Data Platforms

Work Group: Innovative Trial Design Supplemental and External Controls

Date: 21 October 2024



Data Reuse in Regulatory Submissions: The Role of Data Platforms

Abstract

The increasing availability of prior clinical trial data (pCTD) and real-world data (RWD) presents valuable opportunities for using external and supplemental control arms in regulatory submissions. These approaches can reduce patient recruitment needs, accelerate study timelines, and improve trial feasibility, particularly for rare diseases. To realize these benefits and provide confidence in the accuracy of the evidence produced using pCTD/RWD, submitting sponsors and regulators need to be assured of the quality of the data.

Data platforms are an integral part of the ecosystem, enabling trial sponsors to access and use patient data from real-world settings and prior clinical trials. This paper recommends specific steps data platforms can take to support sponsors in meeting health authority expectations.

The data platform practices recommended in the paper are informed by sources that include FDA guidance documents, the HMA-EMA Data Quality Framework, and CRDSA's Standard for Sharing Clinical Study Data. The recommendations span the data lifecycle, including collection, documentation, transparency, data format, and quality assurance (audit/traceability).

By outlining clear expectations to support the regulatory use of supplemental and external controls, the paper highlights the central role played by the data platform. The paper calls for continued collaboration among sponsors, regulators, and data platforms to achieve broad regulatory acceptance of supplemental and externally controlled clinical trials, ultimately improving patient outcomes.

Check for Updates »

About CRDSA

CRDSA is a multi-stakeholder consortium that serves the clinical research data ecosystem.

Our mission is to accelerate the discovery and delivery of lifesaving and life-changing therapies to patients by expanding the research value of patient data from the clinical development process, academic research, and real-world settings. Broad access to these data has the power to transform the research process and improve human health. To find out more please visit **crdsalliance.org**.

Contents

1. Introduction

The availability of individual patient data (IPD) from prior clinical trials (prior clinical trial data, pCTD) and data collected from real-world settings (real-world data, RWD) provides opportunities to use such data in the evaluation of investigational medicines. These types of data can be used to supplement control arms (supplemental control) or provide all the data for a control arm (external control). These uses provide important potential benefits, including:

- Maximizing the scientific use of data contributed by patients
- Decreasing the number of patients randomized to the control arm while not compromising the statistical power of the study — particularly for serious or life-threatening diseases, this may help strengthen the ethical justification for studies and encourage patient participation, as patients are more likely to receive the investigational treatment
- Reducing the time taken to conduct studies by recruiting fewer patients and thereby delivering new medicines more quickly and efficiently
- Improving the feasibility of studies, particularly for rare diseases, where patient recruitment into traditional clinical trials may take a prohibitive amount of time
- Improving trial designs by using the information to better understand how control/standard of care patients may behave in clinical trials so that appropriate estimands are defined to address the key scientific questions to be answered as part of conducting the clinical trial(s)

To realize these benefits and provide confidence in the accuracy of the evidence produced using pCTD/RWD, submitting sponsors and regulators need to be assured of the quality* of the data. Because data platforms are the major providers of these data, they have an important role. This paper summarizes the role of data platforms, discusses the challenges faced by regulators and sponsors, and provides recommendations for data platforms to help submitting sponsors meet regulatory needs.

This paper's recommendations are informed by several CRDSA-sponsored activities that followed the public comment period for the 2023 FDA draft guidance on the use of external controls [1]. These include a regulatory roundtable summarized in CRDSA's External Controls Situation Analysis (August 2023) [2] and a workshop held in January 2024 that included participants representing multiple providers of both pCTD and RWD.

There are other important aspects related to using pCTD/RWD and meeting regulatory expectations for which the submitting sponsor is directly accountable. These include post–data access processing, selection methodologies for pCTD/RWD to minimize or eliminate selection bias, and statistical methods to determine and adjust for sources of bias. These aspects are not directly relevant for data platforms and, therefore, are not included in this paper.

* Data quality is used here as a broad term that encompasses concepts of data integrity including accuracy, consistency, and reliability of data throughout its lifecycle.

2. Health Authority Expectations

In evaluating a new investigational product, health authorities need to be able to characterize, assess, and assure data quality for regulatory decision-making whether data are generated through the clinical trial process or through construction of a supplemental or external control (SEC) arm using pCTD/RWD. This data quality assessment underpins the review of studies by regulators, which may include reproducing analyses reported by the submitting sponsor to ensure the robustness and legitimacy of the conclusions drawn from the data.

The data quality assessment considers whether the data are fit for the purpose of regulatory decision-making. It includes:

- The processes and systems through which data are generated, collected, processed, and made available, which are referred to as foundational elements in the European Medicines Agency (EMA) Data Quality Framework. Examples include the use of certified software systems to collect and process data; the presence of processes, training, and audit processes to ensure data are properly recorded and documented; and the validation and verifiability of data processing steps [3].
- Characteristics inherent in the data. This category includes aspects related to the intrinsic nature of the data (e.g., clinical trial data and claims data) and aspects related to the specific question the data is being used to address [3].

These general principles and concepts related to data quality are largely similar across different regulatory bodies, including the U.S. Food and Drug Administration (FDA), although the precise terminology and definitions may differ. This paper uses the terms data reliability and data relevance as used in FDA guidance [4] when discussing data quality (see Table 1).

Table 1: Data Quality Terms

Term	Description [4]
Data reliability	Includes accuracy, completeness, provenance, and traceability
Data relevance	Includes the availability of key data elements (exposure, outcomes, covariates) and sufficient numbers of representative patients for the study

Traditional clinical trials are conducted according to ICH GCP [5] requirements, with data quality assured through well-established systems for data collection, management, and governance. In evaluating a regulatory submission that includes an SEC arm, the health authority must still be assured that the pCTD/RWD used meets appropriate regulatory standards. In addition, health authorities must also be able to ensure that appropriate audit and inspection processes are in place to facilitate regulatory oversight and assurance of data integrity.

It should be noted that while health authorities generally support and encourage use of these new sources of information [3] [6], incorporation into the regulatory pathway represents a new construct and requires an evolution of the traditional RCT evaluation process to assess the quality and accuracy of these data-at-scale. A good example of this mindset shift is the HMA-EMA Joint Big Data Task Force, which views establishing an EU data framework for data quality and representativeness as a "critical element for realizing the full potential of (big) data and driving regulatory decisions [3]."

3. Submitting Sponsor Challenges

The health authority expectations described above for using pCTD/RWD create several challenges for submitting sponsors when using these data to construct an SEC arm for regulatory submission.

Demonstrating data quality to regulators. Data collection and processing that takes place before the submitting sponsor accesses the data are subject to third-party quality management systems, processes, and transformations. Therefore, submitting sponsors may not have ready access to the information needed to adequately assess whether these data are fit for purpose and demonstrate the data quality expected by regulators in a new submission.

Ability to download patient-level data. Some data platforms limit secondary use dataset access to on-platform or remote access. This can be problematic for sponsors contemplating use in a regulatory submission, where secondary use IPD may be required by the health authority.

Data format. pCTD generated by biopharmaceutical sponsors is delivered in the industry-standard CDISC format (Study Data Tabulation Model, SDTM). However, RWD is not routinely delivered for use in SDTM, but may be provided in a variety of formats, which differ between data platform providers. Mapping these disparate data formats to SDTM suitable for regulatory submission requires a considerable commitment of sponsor resources, potentially duplicated across multiple sponsors using the same RWD resource.

Audit and inspection. To comply with applicable privacy laws and regulations, patient-level data provided by third parties is anonymized†, and the submitting sponsor will not have access to the original source data. Further, there may be contractual, ethical, practical, and regulatory limitations (e.g., legal basis) that render sponsor, health authority, and data platform access to the secondary use source unfeasible or impossible.

Although these challenges relate to meeting regulatory expectations, a sponsor using pCTD/RWD for a control arm needs to be assured of data quality irrespective of whether the clinical study is ultimately submitted to regulators. For example, an incorrect decision or statistical inference leading to discontinuation of medicine development because of data quality issues would not be in the best interests of patients or the sponsor company.

† Anonymization is used throughout this paper to broadly include all forms of privacy protection, recognizing that regulatory bodies may use other terms (e.g., de-identification) and methods.

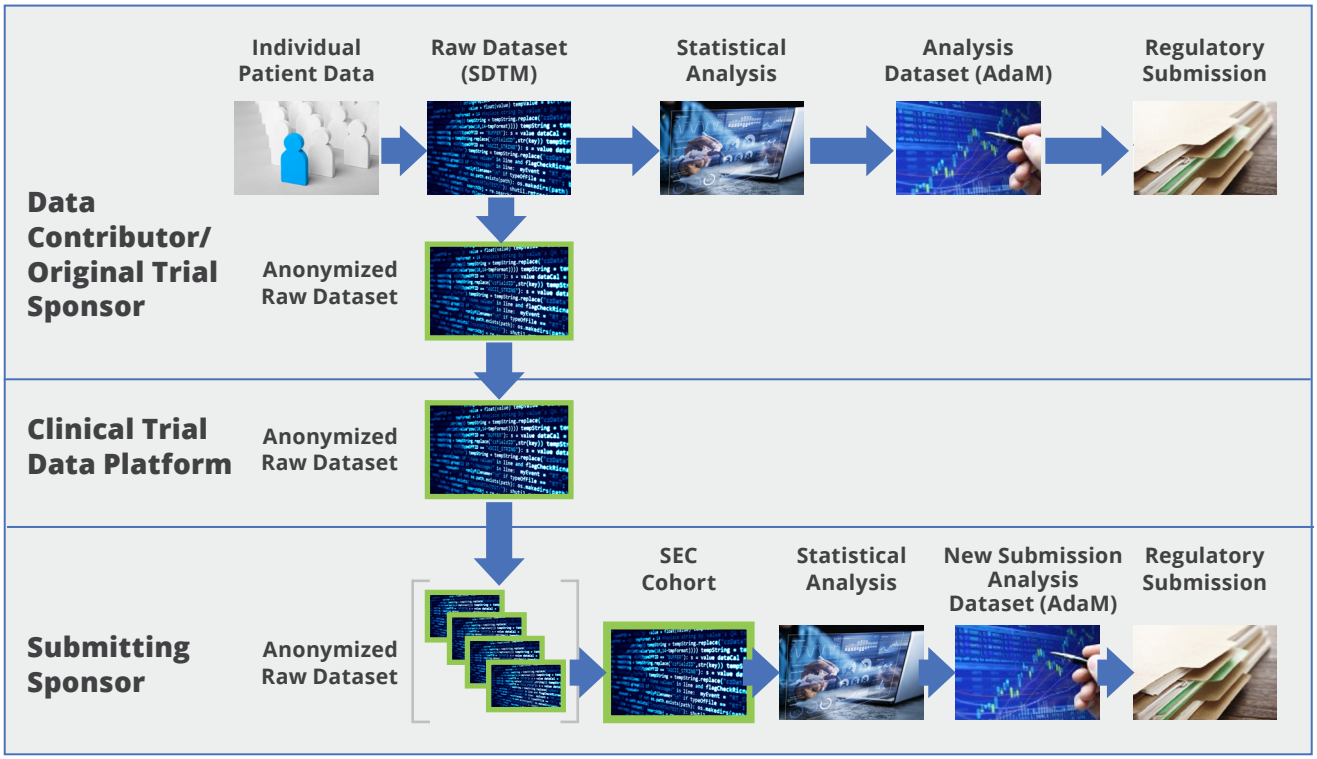
4. The Role of Data Platforms

Data platforms are an integral part of the ecosystem, enabling the access and use of secondary data by trial sponsors. There are, however, important differences in how pCTD and RWD are collected, prepared, and accessed for secondary data use.

Clinical Trial Data

Data platforms hosting clinical trial data provide a way for sponsors to request and access pCTD that has been anonymized for data sharing by data contributors. These data can then be used for control arms and submitted as part of a regulatory submission (see Figure 1). Data platforms act as neutral repositories of contributed trial data and, in some cases, may perform additional transformations to prepare data for secondary use (such as creating harmonized cohort pools with data from multiple trials).

Figure 1: Clinical Trial Data: Platforms Provide Access to Anonymized Datasets Processed by the Data Contributor

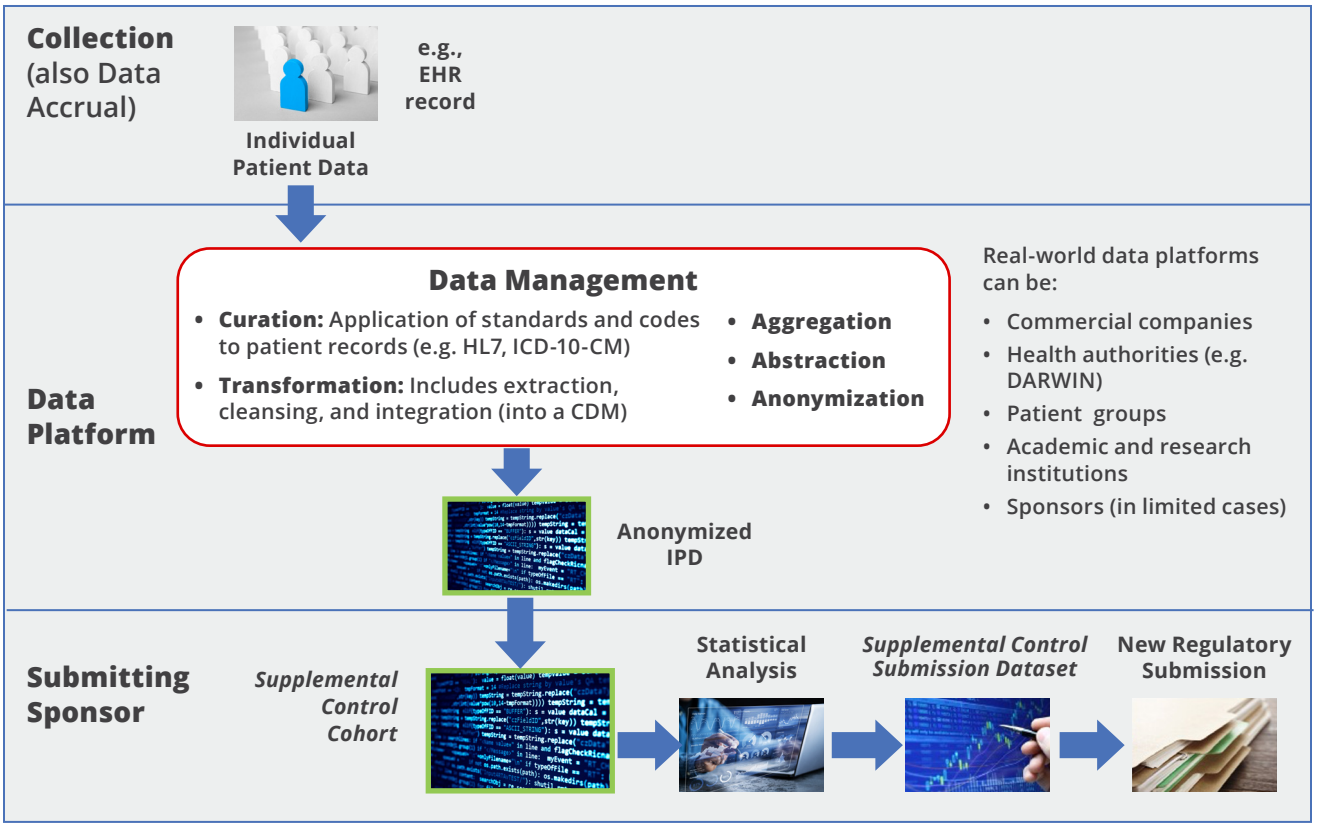


Real-World Data

Preparing RWD for secondary data use involves a series of processing steps that are typically performed by the data platform (see Figure 2). Although the data lifecycle may vary depending on the type of data and setting (e.g., healthcare settings such as pharmacies, clinics, emergency departments, and hospitals), in general, the lifecycle involves multiple phases: data accrual from the original source data, curation of data to the clinical data repository, and transformation and anonymization of data. Submitting sponsors can access these data and use the data in control arms when producing a study-specific dataset for analysis (see Figure 2) [4].

It should also be noted that access to RWD can vary depending on both source and geographic location. For example, in the United States, RWD is generally licensed through commercial data platforms, while in the European Union, RWD access for regulatory decision making is facilitated through EMA's Data Analysis and Real World Interrogation Network (DARWIN) or other EMA-sponsored data pathways. [7]

Figure 2: RWD: Platforms Process Data for Secondary Use (see Definitions)



5. Data Platform Recommendations

The following recommended practices are designed to help ensure that submitting sponsors are able to meet regulatory expectations while adhering to the privacy and source data access constraints of using data provided by data platforms (considered "third parties" by health authorities such as the FDA).

5.1 Data Collection

These recommendations address the processes and controls needed to generate confidence in the source (original) data used to generate the secondary use data.

5.1(a) Data Collection: pCTD

Where data platforms create pooled data containing IPD from multiple trials, the platform should either:

- **Provide a dataset where only clinical trials that have been submitted to health authorities for regulatory review are included, or**
- **Enable sponsors to identify and extract for their analysis clinical trials that have been submitted to health authorities for regulatory review.**

In clinical trials intended for regulatory review, source data generation and collection will be carried out in accordance with health authority requirements and generally accepted practices (e.g., ICH GCP [5]). Data collection and data management processes are documented in sponsor SOPs and the study protocol for the specific study. Robust quality management processes are also in place to assure the accuracy and completeness of the data.

Sponsors using pCTD to construct an SEC control arm should, whenever possible, use data from trials that had been submitted to health authorities for regulatory review. Where data are made available for an SEC from a clinical trial that has not been submitted to health authorities, the submitting sponsor who intends to use the data may need to discuss the specifics with health authorities or use the data for exploratory rather than pivotal studies.

Data platforms hosting clinical trial data should support sponsors by ensuring trials that were subject to health authority review can be readily identified and, when needed (such as when data from multiple trials are combined in a pooled cohort), ensuring that sponsors can readily access or extract the previously reviewed trials.

5.1(b) Data Collection: RWD

The process for capturing, entering, or collecting data (e.g., data curation from the EHR) is subject to clearly defined and documented SOPs, with quality control procedures documented and implemented.

RWD is typically collected outside of a protocol-driven research setting and is not subject to harmonized requirements equivalent to ICH GCP [5]. As a result, there will be more variability in how data owners and custodians collect RWD. It is therefore important that the processes for collecting data (e.g., data curation from the EHR) are clearly defined and documented in SOPs and data manuals that are included in training for relevant personnel. In addition, as with traditional clinical trials, there should be quality control processes to assure the accuracy and completeness of the data.

5.2 Data Documentation

These recommendations are intended to foster consistency and transparency sufficient to support the regulatory use of secondary data.

5.2(a) Data Documentation: pCTD

Data platforms should ensure that contributed trials include sufficient data and supporting documentation to facilitate secondary research and allow health authorities to assess data relevance and data reliability. The CRDSA Standard for Sharing Clinical Study Data [8] introduces requirements to support this recommendation.

Data platforms should ensure that datasets and supporting documentation are available to data users such that they can readily be used in a new health authority submission (e.g., through data download or other method acceptable to health authorities).

The CRDSA Standard for Sharing Clinical Study Data [8] requires study documents to be shared with anonymized datasets. This is to enable assessment of data relevance for the scientific question being asked and enable researchers to use the data provided. These documents include:

- Published information and metadata
- Latest study protocol
- Dataset specification
- Data dictionary
- Annotated case report form
- Statistical analysis plan
- Clinical study report
- Encoding information

In addition, to enable researchers to understand how the data has been anonymized and to enable assessment of data relevance and reliability, the CRDSA standard requires a data transformation report to be provided with the datasets to detail at a domain and variable level how the data has been changed to protect privacy [8].

Certain health authorities, including the FDA, require the patient-level data used to construct an SEC to be submitted for regulatory review. Further, the supporting pCTD trial documentation gives regulators important context in the review process. Therefore, there should be limited contractual or technical restrictions on using these data and documents in a new regulatory submission.

5.2(b) Data Documentation: RWD

Where RWD is to be used in a regulatory submission, data platforms should support sponsors in complying with FDA guidance (Dec 2023), Data Standards for Drug and Biological Product Submissions Containing Real-World Data [9], including:

- **Process documentation: The end-to-end data collection, curation, and management process is documented and available to end users for review and, when required, submission to health authorities. Documentation should include process and quality control procedures for all data changes from the source data system to secondary use data availability for platform end users.**
- **Dataset-level transformation documentation: RWD datasets should include documentation at the domain and variable level for any changes, alterations, deletions, or transformations from the original value collected at source to the secondary use dataset (including deletions, additions, and alterations).**
- **Data format: When RWD is to be used in a regulatory submission, data platforms should, at the request of the sponsor, be able to facilitate provision of datasets in SDTM [10] format.**

Process Documentation

For traditional clinical trials, processes applied in transforming data from the source data into analysis-ready datasets is described in process documents (e.g., SOPs) that meet health authority requirements.

Similarly, for RWD, data management processes applied to the data (e.g., curation, transformation, aggregation, abstraction, anonymization — see Figure 2) are to be documented and made available to submitting sponsors who can review the information so they have confidence in the data processes and can incorporate relevant information into their submission.

Dataset-level Transformation Documentation

Source data from a specific traditional clinical trial is managed and transformed according to a data management plan for that trial. This provides documentation for creating SDTM and ADaM [10] [11] datasets.

Similarly, where RWD is used in an SEC, health authorities require sponsors to provide the patient-level data in SDTM format, along with supporting documentation on any domain and variable level changes, alterations, deletions, or transformations from the source data to the dataset provided to the submitting sponsor. This is important for data provenance and traceability, allowing data to be traced back to its source and to understand any modifications made along the way.

Data Format

To support sponsors, RWD data platforms should facilitate the provision of datasets to be used in regulatory submissions in SDTM format and document the data mapping approach and any challenges encountered. This documentation is to include the anticipated impact of data mapping and definitions of data elements derived from RWD sources [9]. Supplying this documentation with the datasets provided enables the information to be assessed by submitting sponsors and incorporated into submissions for health authority review.

The authors recognize that the provision of RWD in SDTM format, with complete transformation detail, is not currently a general practice. Further, it is recognized that facilitating the provision of "regulatory ready" datasets may require additional data platform resources and associated costs. However, because platforms supply data to multiple sponsors, facilitating the provision of data in the standard regulatory format (before sponsor analysis) can help reduce the variability inherent in different sponsors creating multiple SDTM datasets from the same source. This ensures consistent interpretation of the data, promotes interoperability, and serves to establish the data platform as the single source of truth.

5.3 Quality Assurance

Health authorities have the responsibility to evaluate data quality and to ensure the veracity and legitimacy of data submitted for regulatory review. The following recommendations support that regulatory responsibility within the context of secondary use data. For all regulatory submissions, it is critical for sponsors to have clarity regarding the potential audit risk. In the secondary data use ecosystem, data platforms play a fundamental role in generating both sponsor and health authority confidence in the data they provide.

5.3(a) Quality Assurance: Data Collection – pCTD

Not applicable for prior clinical trial data collected under industry guidelines.

Where data for an SEC is provided from a prior clinical trial that has been submitted to health authorities for review, the study has already been subject to health authority scrutiny from a data quality perspective — including possible audit and inspection of source data and documents. Therefore, the submitting sponsor and health authorities can be assured of data reliability without the need for health authority audit of the source clinical trial data.

5.3(b) Quality Assurance: Data Collection – RWD

RWD data platforms should facilitate audits and inspections that allow end users (sponsors) and health authorities to ensure adherence to collection and data management SOPs.

Because RWD is typically collected outside a clinical research setting (see 5.1b), RWD data platforms should facilitate audits and inspections that allow sponsors and health authorities to ensure that appropriate and adequate data collection processes and collection quality assurance SOPs are in place to ensure data validity and completeness at the point of data collection (e.g., data curation from the EHR system).

5.3(c) Quality Assurance: Data Platform Transformations – pCTD

If the data platform has performed data transformations on clinical trial data, the data (prior to any transformations) should be made available for audit and inspection by health authorities to assess data reliability.

Some data platforms transform anonymized clinical trial data provided by data contributors. For example, they may pool clinical trials for a specific disease. Where this involves transformation of the data provided, data platforms should make available the data *prior to any transformations performed by the data platform* for sponsor and health authority inspection and audit so they can assess data reliability.

5.3(d) Quality Assurance: Data Platform Transformations – RWD

Data platforms should make source real-world data available for health authority inspection and audit.

As shown in Figure 2, data platforms typically perform comprehensive data management operations in preparing RWD for secondary use. RWD data platforms should allow and facilitate health authority audits and inspections of the source data held by the data platform prior to any alterations or transformations performed by either the data platform or, as applicable, the submitting sponsor. This is consistent with the FDA guidance (Aug 2023), Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products [12] which states:

> *"If certain RWD are owned and controlled by other entities, sponsors should have agreements in place with those entities to ensure that relevant patient-level data can be provided to FDA and that source data necessary to verify the RWD are made available for inspection as applicable."*

5.3(e) Quality Assurance – Secondary Use Submission Dataset

Data platforms should facilitate, through data access or other comparable method, health authority audit of the secondary use data used in the regulatory submission such that health authorities are able to verify that the data used to create the study analysis dataset is the same as that provided by the platform to the sponsor.

To assure data reliability for a new regulatory submission, it is important that health authorities can verify that the data used by the submitting sponsor has been faithfully derived from data provided by the data platform. Several approaches may be available to enabling data verification, including providing health authorities access to the secondary use data provided to submitting sponsors.

Table 2: Data Platform Summary Recommendations

Prior Clinical Trial Data (pCTD)	Real World Data (RWD)
Data Collection	
Where data platforms create pooled data containing IPD from multiple clinical trials, the platform should either: • Provide a dataset where only clinical trials that have been submitted to health authorities for regulatory review are included, or • Enable sponsors to identify and extract for their analysis clinical trials that have been submitted to health authorities for regulatory review.	The process for capturing, entering, or collecting data (e.g., data curation from the EHR) is subject to clearly defined and documented SOPs, with quality control procedures documented and implemented.
Data Documentation	
Data platforms should ensure that contributed trials include sufficient data and supporting documentation to facilitate secondary research and allow health authorities to assess data relevance and data reliability. The CRDSA Standard for Sharing Clinical Study Data [8] introduces requirements to support this recommendation. Data platforms should ensure that datasets and supporting documentation are available to data users such that they can readily be used in a new health authority submission (e.g., through data download or other method acceptable to health authorities).	Where RWD is to be used in a regulatory submission, data platforms should support sponsors in complying with FDA guidance (Dec 2023) Data Standards for Drug and Biological Product Submissions Containing Real-World Data [9] including: • Process documentation: The end-to-end data collection, curation, and management process is documented and available to end users for review and, when required, submission to health authorities. Documentation should include process and quality control procedures for all data changes from the source data system to secondary use data availability for platform end users • Dataset-level transformation documentation: RWD datasets should include documentation at the domain and variable level of any changes, alterations, deletions, or transformations from the original value collected at source to the secondary use dataset (including deletions, additions, and alterations) • Data format: When RWD is to be used in a regulatory submission, data platforms should, at the request of the sponsor, be able to facilitate provision of datasets in SDTM format.
Quality Assurance	
Data Collection	
Not applicable for prior clinical trial data collected under industry guidelines.	Data platforms should facilitate audits and inspections that allow end users (sponsors) and health authorities to ensure adherence to collection and data management SOPs.
Data Platform Transformations	
If the data platform has performed data transformations on clinical trial data, the data (prior to any transformations) should be made available for audit and inspection by health authorities to assess data reliability.	• Data platforms should make source data available for health authority inspection and audit consistent with the FDA guidance (Aug 2023): Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products [12].
Secondary Use Submission Dataset	
Data platforms should facilitate, through data access or other comparable method, health authority audit of the secondary use data used in the regulatory submission such that health authorities are able to verify that the data used to create the study analysis dataset is the same as that provided by the platform to the sponsor.	

6. Benefits and Challenges for Data Platforms

By implementing these principles data platforms can:

- Enhance the quality and reliability of pCTD/RWD, increasing its value for secondary research purposes
- Facilitate more efficient and effective use of pCTD/RWD by providing comprehensive documentation and metadata
- Foster trust and confidence among submitting sponsors and regulators in the platform's data sharing practices
- Help streamline the regulatory submission process by ensuring that shared data meets regulatory expectations

However, it is recognized that implementation of these recommendations by data platforms may require additional resources to further ensure and document data quality. Nonetheless, the potential benefit for improving patient outcomes is substantial; as the demand for secondary use pCTD/RWD increases, platforms that are able to demonstrate and document data quality will be well positioned to meet the needs of submitting sponsors and regulators.

7. Conclusion

The use of pCTD and/or RWD as supplemental and external controls provides opportunities for medicines to be developed to meet the needs of patients more quickly. There are challenges for submitting sponsors in meeting regulatory expectations when using these kinds of data because of privacy and third-party source data access constraints. The recommendations for data platforms provided in this paper will help sponsors meet these regulatory expectations within this framework (see Table 3). We call on sponsors, regulators, and data platforms to further develop these recommendations and provide implementation best practices that can be applied by data platforms.

Table 3: Data Platform Recommendations Support Regulatory Expectations

Regulatory Expectation	Recommendation
The submitting sponsor has responsibility for demonstrating data quality (from data generation/collection through to data submission).	Data Collection – pCTD 5.1(a) Data Collection – RWD 5.1(b) Data Documentation – pCTD 5.2(a) Data Documentation – RWD 5.2(b)
Patient-level data is submitted to regulators in SDTM format with sufficient data transformation transparency, to support regulatory review.	Data Documentation – pCTD 5.2(a) Data Documentation – RWD 5.2(b)
Regulators can access appropriate and sufficient data and documents for inspections/audits.	Quality Assurance – pCTD 5.3(a), 5.3(c), 5.3(e) Quality Assurance – RWD 5.3(b), 5.3(d), 5.3(e)

There are other important aspects related to using pCTD/RWD for SEC and meeting regulatory expectations, which are not addressed in this paper. These include selection methodologies for pCTD/RWD to minimize or eliminate selection bias, statistical methods to determine and adjust for sources of bias, guidelines on the appropriate power for supplemental and fully external control arms, and development of methods of multimodal analysis methods and best practices when constructing SECs with IPD from both pCTD and RWD.

We encourage and support a continued effort to further develop consensus approaches that will provide a comprehensive road map for how submitting sponsors can effectively use pCTD and RWD in regulatory submissions and thereby meet the needs of patients.

8. Appendix: Regulatory (FDA) Expectations for Clinical Trials using pCTD/RWD

Regulatory Expectation	FDA Guidance (examples)	Reference
The submitting sponsor has responsibility for demonstrating data quality.	"Sponsors should demonstrate the appropriateness of the proposed data source(s) to address specific hypotheses and research questions."	[13]
	"Additionally, sponsors should ensure that RWD were collected using good data management practices and are sufficiently robust."	[4]
	"Sponsors should enable and maintain audit trails of data, starting from extracting RWD sources through maintenance and retention of dataset(s). This process should include the tracking of user access, data changes, changes to the protocol, and analyses performed."	[12]
	"When considering RWD sources for regulatory purposes, sponsors should consider the methods and systems used to help ensure sufficient data quality, including any data quality assurance plans and	[4]
Patient-level data is submitted in SDTM and ADaM formats.	"If sponsors do not own the data used for the external control arm, they should structure their agreements with the data owners to ensure that patient-level data can be provided to FDA in support of the marketing application."	[1]
	"When the RWD source is not owned by the sponsor, the sponsor should attempt to obtain participant-level data for each participant. If not available, the sponsor should define the entity(ies) which do have access/permission for data entry, quality assurance, storage, aggregation or other linkage, and assessment of traceability from data entry to dataset, as applicable. Sponsors should consider the level of access which could be shared with FDA and the potential for third parties to provide participant-level data directly to FDA."	[4]

Regulatory Expectation	FDA Guidance (examples)	Reference
Patient-level data is submitted in SDTM and ADaM formats.	"Sponsors must ensure that they are able to submit patient-level data for any RWD that have been analyzed as part of the clinical study included in a marketing application when required under 21 CFR 314.50 and 601.2."	[12]
	"With adequate documentation of the conformance methods used and their rationale, study data derived from RWD can be transformed into SDTM and ADaM datasets and submitted to FDA in an applicable submission."	[9]
Regulators can access source data and documents for inspections and audits.	"Sponsors should also ensure that FDA has access to Source Documents and Source Data for the external control arm as part of an FDA inspection or upon request."	[1]
	"If certain RWD are owned and controlled by other entities, sponsors should have agreements in place with those entities to ensure that relevant patient-level data can be provided to FDA and that source data necessary to verify the RWD are made available for inspection as applicable."	[12]

9. Glossary

Term	Definition
Abstraction	The process of simplifying complex datasets by focusing on relevant details and concealing unnecessary intricacies. This involves creating a higher-level representation of data that captures essential features while omitting nonessential details.
Anonymization	The process of modifying or removing personally identifiable information from a dataset to protect individual privacy while retaining the utility of the data for analysis.
Collection	The systematic process of gathering, recording, and assembling information or observations to build a dataset.
Curation	Application of standards (e.g., HL7, ICD-10-CM) to source data; for example, the application of codes to adverse events, disease staging, the progression of disease, and other medical and clinical concepts in an EHR.
Data accrual	The process by which the data was collected
Data aggregation	The process of combining and summarizing individual data points into a consolidated and more manageable form. This involves grouping, averaging, or otherwise transforming raw data to derive higher-level insights, often to analyze trends, patterns, or statistical measures.
Data cleansing	Data cleansing (sometimes referred to as data scrubbing) is the process of correcting or removing inaccurate, improperly formatted, or duplicate data or records from a database. The data requiring correction/removal is sometimes referred to as "dirty data." Data cleansing is an essential task for preserving data quality
Data integrity	The completeness, consistency, and accuracy of data
Data provenance	An audit trail that “accounts for the origin of a piece of data (in a database, document, or repository) together with an explanation of how and why it got to the present place”
Data relevance	Refers to the availability of key data elements (exposure, outcomes, covariates) and sufficient numbers of representative patients for the study
Data reliability	Refers to the accuracy, completeness, provenance, and traceability for data accrual, curation, and transformation into the final study-specific dataset

Term	Definition
Data standardization	The systematic process of converting data into a uniform format or structure to ensure consistency and compatibility across different datasets. This involves establishing common rules, conventions, or units of measurement for variables, making the data comparable and interoperable.
Data traceability	Permits an understanding of the relationships between the analysis results (tables, listings, and figures in the study report), analysis datasets, tabulation datasets, and source data
Data transformation	Includes data extraction, cleansing, and integration (e.g., into a CDM)
Data validation	The process of establishing that a method is sound or that data are correctly measured, usually according to a reference standard
Derived variable	A calculated or transformed measure generated from one or more existing variables in a dataset
Electronic health record (EHR)	An individual patient record contained within an electronic health system. A typical individual EHR may include a patient's medical history, diagnoses, treatment plans, immunization dates, allergies, radiology images, pharmacy records, and laboratory and test results.
External control	A clinical trial control arm that contains data entirely from previously conducted clinical trials and/or real-world data (RWD)
Medical claims data	The compilation of information from medical claims that health care providers submit to insurers to receive payment for treatments and other interventions. Medical claims data use standardized medical codes, such as the World Health Organization's International Classification of Diseases (ICD-CM) diagnosis codes, to identify diagnoses and treatments.
Real-world data (RWD)	Data relating to patient health status or the delivery of health care routinely collected from a variety of sources
Real-world evidence (RWE)	Clinical evidence regarding the usage and potential benefits or risks of a medical product derived from analysis of RWD
Secondary use	Use of data for a purpose other than the purpose for which it was originally collected and processed

Term	Definition
Secondary use data	The datasets and documentation on the data platform available for secondary use (after data preparation by the data contributor and/or data platform)
Source data	Data and documents as collected for primary use (e.g., in a prior clinical trial or an electronic health record)
Source data verification	A process used in clinical trials to ensure the accuracy, reliability, and integrity of data collected during the study. It involves the comparison of data recorded on case report forms (CRFs) or electronic data capture systems with the original source documents from the participating sites.
Submission dataset	The final study analytic dataset submitted by the sponsor for regulatory review
Submitting sponsor	Sponsor who uses secondary use datasets for SEC for clinical studies submitted to regulatory authorities
Supplemental control	A clinical trial control arm where a portion of the control arm includes data from previously conducted clinical trials and/or real-world data (RWD)
Validation	The process of checking the data within a dataset to ensure it meets certain quality criteria, such as falling within an expected range
Verification	The process of comparing the data in the dataset against the original source records to ensure accuracy and completeness

10. References

[1] U.S. Food and Drug Administration "Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products: Guidance for Industry: Draft Guidance," February 2023. [Online]. Available: https://www.fda.gov/media/164960/download

[2] Clinical Research Data Sharing Alliance, "FDA Draft External Controls Guidance: Situation Analysis and Recommendations," August 21, 2023. [Online]. Available: https://crdsalliance.org/crdsa_resources/fda-draft-external-controls-guidance/

[3] European Medicines Agency, "Data Quality Framework for EU medicines regulation," October 30, 2023. [Online]. Available: https://www.ema.europa.eu/en/documents/regulatory-procedural-guideline/data-quality-framework-eu-medicines-regulation_en.pdf

[4[U.S. Food and Drug Administration, "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products: Guidance for Industry," July 2024. [Online]. Available: https://www.fda.gov/media/152503/download

[5] International Council for Harmonization to Technical Requirements for Pharmaceuticals for Human Use, "ICH Harmonized Guideline Integrated Addendum to ICH E6(R1): Guideline for Good Clinical Practice E6(R2)," November 2016. [Online]. Available: https://database.ich.org/sites/default/files/E6_R2_Addendum.pdf

[6] U.S. Food and Drug Administration, "21st Century Cures Act," FDA.gov. Accessed: Oct. 9, 2024. [Online]. Available: https://www.fda.gov/regulatory-information/selected-amendments-fdc-act/21st-century-cures-act

[7] Heads of Medicines Agencies, European Medicines Agency, "Real-world evidence provided by EMA: Support for regulatory decision-making," April 10, 2024. [Online]. Available: https://www.ema.europa.eu/en/documents/other/guidance-real-world-evidence-provided-ema-support-regulatory-decision-making_en.pdf

[8] Clinical Research Data Sharing Alliance, "CRDSA Std 1001: Standard for Sharing Clinical Study Data," September 25, 2024. doi: 10.70496/jcq17x. [Online]. Available: https://crdsalliance.org/crdsa_resources/crdsa-std-1001-standard-for-sharing-clinical-study-data/

[9] U.S. Food and Drug Administration, "Data Standards for Drug and Biological Product Submissions Containing Real-World Data: Guidance for Industry," December 2023. [Online]. Available: https://www.fda.gov/media/153341/download

[10] CDISC, "SDTM," CDISC.org. Accessed: Oct. 9, 2024. [Online]. Available: https://www.cdisc.org/standards/foundational/sdtm

[11] CDISC, "ADaM," CDISC.org. Accessed: Oct. 9, 2024. [Online]. Available: https://www.cdisc.org/standards/foundational/adam

[12] U.S. Food and Drug Administration, "Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products: Guidance for Industry," August 2023. [Online]. Available: https://www.fda.gov/media/171667/download

[13] U.S. Food and Drug Administration, "Real-World Evidence: Considerations Regarding Non-Interventional Studies for Drug and Biological Products: Guidance for Industry: Draft Guidance," March 2024. [Online]. https://www.fda.gov/media/177128/download