Paper RE05

Submitting RWD: Where We Are and Where We Are Going

Ingeborg Holt, Orizaba Solutions, USA

Sarah Ferko, IBM Consulting, USA

Jeff Abolafia, P21, Blue Bell, PA, USA

ABSTRACT

The rapid increase in the use of non-interventional study data from real world sources to support marketing applications has posed numerous challenges for both regulatory reviewers and sponsors. This presentation examines the current regulatory environment and proposed standards for submitting Real World Data (RWD). We review responses to the FDA's request for comments in the Federal Register Notice (FRN), Docket No. FDA–2025–N–0287 titled "Exploration of Health Level Seven Fast Healthcare Interoperability Resources (FHIR) for Use in Study Data Created From Real-World Data Sources for Submission to the Food and Drug Administration" to evaluate the expected impact of two proposed options for submitting RWD, expanding CDISC standards to accommodate evidence derived from RWD, and submitting RWD using the HL7® FHIR® standard. We also provide short term and long recommendations for submitting randomized controlled trials and RWD study data.

INTRODUCTION

There has been an increase in the number of New Drug Applications (NDAs) and Biologics License Applications (BLAs) that include Real World Data (RWD), data from retrospectively designed observational studies, and that increase is expected to continue. Over twenty years ago when Clinical Data Interchange Standards Consortium (CDISC) standards were initially developed to standardize the data submitted to support marketing applications, NDAs and BLAs were primarily a collection of randomized controlled trials (RCTs). The rise in the use of electronic healthcare records (EHRs) was spurred by regulation including the Health Information Technology for Economic and Clinical Health Act (HITECH) Act and the Affordable Care Act (ACA) which increased the use of EHRs and thus the availability of electronic healthcare data. With the increasing volume of electronic data available from EHRs, claims, and other sources, there has been an increasing push to make use of this data for other purposes, including bringing medical products to market. Consequently, there has been a rapid increase in the trend to include RWD from retrospectively designed non-interventional studies in marketing applications.

While observational studies have been used in the pre-market setting to provide regulatory authorities with evidence of efficacy and/or safety, more widespread and consistent use currently presents numerous challenges for both regulatory reviewers and sponsors. In our previous papers (1, 2, 3), we examined the current regulatory landscape for submitting RWD and the gaps and limitations of using the current data standards for submitting RWD, including the challenges associated with traceability, and offered recommendations for remedying these challenges using the current standards. In this paper, we review the responses from 44 different stakeholders to FDA's request for comment on using FHIR as the data standard for the submission of study data created from RWD to support new NDA and BLA applications. We consider four of the five total questions posed by the FDA and review and summarize the responses. The questions posed include:

1. What challenges do you see for the pharmaceutical industry regarding the current state of submitting clinical study data collected from RWD sources to FDA?

2. What opportunities and/or challenges do you see for the pharmaceutical industry on reaching a future state of clinical study data submissions collected from RWD sources using HL7 FHIR (e.g., business processes, technical considerations)?

3. What are your suggestions on how, from a data standards perspective, FDA might reach a future state of clinical study data submissions collected from RWD sources that aligns with ASTP/ ONC health IT goals for HL7 FHIR-based exchange?

4. Does USCDI version 3 provide enough information for collecting RWD for research purposes? Is there information that USCDI version 3 does not sufficiently address?

We then summarize the advantages and disadvantages of using FHIR as a submission standard for RWD based on the responses and our knowledge evaluating RWD use at FDA for CDER and CBER. Our review and final recommendations are intended to illuminate both advantages and issues associated with using HL7 FHIR, as well as to encourage collaboration and discussion between sponsors, standards development organizations (SDOs), and regulatory authorities to determine the best path forward to enable the most efficient and effective submission of RWD to ensure that safe and effective medical products get into the hands patients as expeditiously as possible.

BACKGROUND

REAL WORLD EVIDENCE

Real World Evidence (RWE) can come from a variety of sources including EHRs, payer administration claims, digital health technologies (DHTs), and patient registries. Sponsors are submitting an increasing amount of RWD to support marketing applications, and this trend is likely to accelerate. Aetion (4) estimates that approximately 78% of approved NDAs and BLAs in 2020 included a study containing RWD to provide evidence of safety and/or efficacy, compared to only 53% in 2019. Further, a study in 2022 estimated that 95% of approved NDAs and BLAs in 2021 included an RWE study (5).

Section 505F(b) of the FD&C Act defines RWE as "*data regarding the usage, or the potential benefits or risks, of a drug derived from sources other than traditional clinical trials*" (21 U.S.C. 355g(b)). In developing its RWE program, FDA believes it is helpful to distinguish between both the sources of RWD, and the evidence derived from that data. Evaluating RWE in the context of regulatory decision-making depends not only on the evaluation of the methodologies used to generate the evidence, but also on the reliability and relevance of the underlying RWD, and these constructs may raise different types of considerations. For the purposes of this framework, FDA defines RWD and RWE as follows (6):

- **Real-World Data (RWD)** are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources; and

- **Real-World Evidence (RWE)** is the clinical evidence about the usage and potential benefits or risks of a medical product derived from analysis of RWD.

For the purposes of this paper, we will use the working definition of RWD as data relating to patient health status and/or the delivery of health care that is not collected under a protocol.

REGULATORY LANDSCAPE

The 21st Century Cures Act (7) was enacted in December 2016, and mandated that the FDA create a framework to evaluate RWE to support marketing applications for new drugs or new indications for approved drugs. Subsequently, in December 2018, the FDA published its

framework for RWE, the *Framework for FDA'S Real World Evidence Program* (6). The framework focused on three areas:

(1) Whether RWD are fit for use;

(2) Whether the study design can provide adequate evidence; and

(3) Whether the study conduct meets regulatory requirements.

The FDA focused on these three areas when evaluating RWD submitted in support of a marketing application. In 2021, the FDA expanded its program by issuing four more guidance documents that provided additional detail, as well as implementation strategies for using RWD to support marketing applications for drugs or biological products. The first guidance, *Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision Making for Drug and Biological Products* (8) focused on the validation of definitions for study design elements and data provenance and quality throughout the study lifecycle. This guidance was finalized in 2023.

In October 2021, the FDA issued a second guidance, *Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry* (9), which outlined the data standards required when submitting RWD in support of a marketing application; a final version of this guidance was published in 2023. The primary takeaway from this guidance is that RWD must be submitted using the standards documented in the FDA Data Standards Catalog (10). To date, this means that RWD *must* be submitted using CDISC standards; thus, sponsors are tasked with converting real-world datasets into CDISC format when submitting this information in support of a marketing application. This guidance also provides advice on mapping RWD in CDISC format and considerations for required documentation. The agency acknowledges that its current catalog of standards does not necessarily reflect data derived from real-world sources and is actively considering other standards and recommendations for mapping. With respect to RWD, the guidance advises that sponsors should confirm that they are able to submit patient-level data for any RWD used in studies in support of a marketing application required under an NDA and/or BLA, however, this decision may be reversed under the FDA Commissioner Makary (11). Additional RWD/RWE guidance documents published by FDA are not summarized in this paper but are publicly available on the FDA RWE site: https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

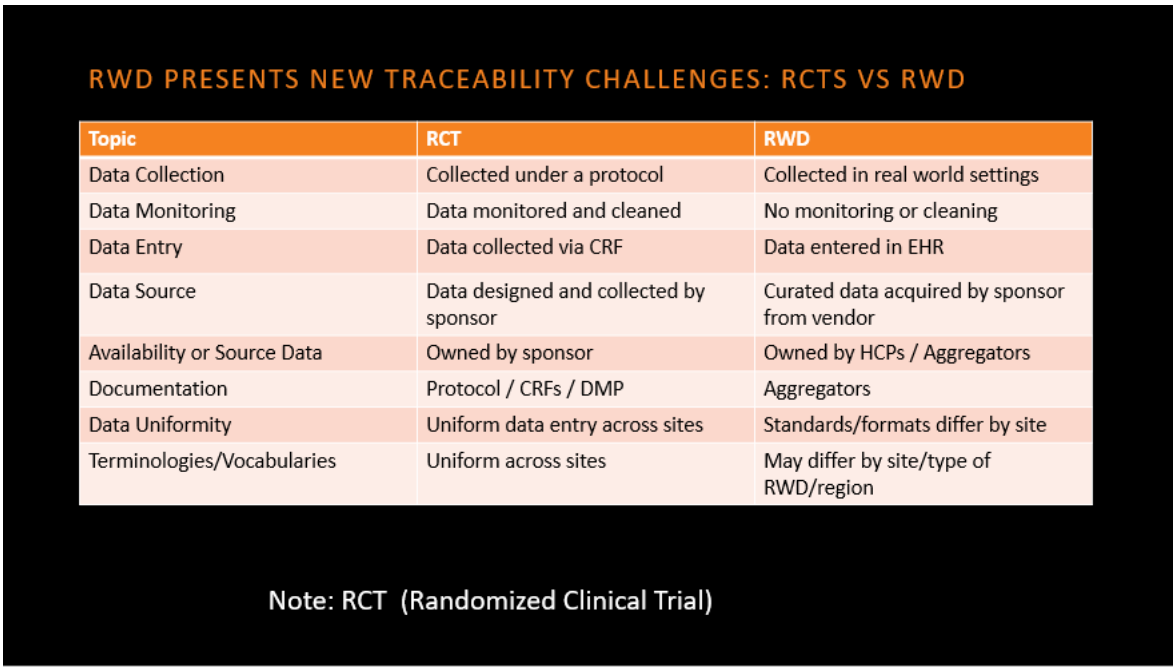
CURRENT STATE

Challenges Submitting RWD Using CDISC Standards

The landscape for submitting RWD is still evolving. The FDA guidance documents cited above indicate that the evidentiary standards, data standards, and documentation required for submitting RWD are the same as for submitting RCT data, which presents numerous challenges for sponsors. In our previous papers, we examined the gaps and limitations for submitting RWD using the current data standards as well as the challenges of providing the traceability required for regulatory review. These challenges are summarized in Figure 1. Note that most of these challenges stem from the fact that RWD, by definition, is not collected under the supervision of a protocol and by research study staff. In addition, the business cases for using RWD and RCTs differ. RWD is not designed or collected for research purposes. As a result, real world databases are organized and configured in a way that makes sense to health care providers, not researchers. Further, these databases utilize different data standards and terminologies that

contain different concepts and coding systems compared to those defined within CDISC and used in regulatory review.

Figure 1. Traceability Challenges: RCTs vs RWD



Current Standards Considered for RWD and RWE

CDISC and FHIR

CDISC currently presents numerous challenges for submitting clinical data to both sponsors and reviewers. In previous papers, we reviewed standards designed to represent RWD including HL7 FHIR, OMOP, PCORNet, and Sentinel. The FDA also evaluated these standards for submitting RWD. Subsequently, FHIR has emerged as a possible standard for submitting data derived from RWD. In 2025, FDA issued a FRN to gather comments from industry on submitting RWD using the FHIR standard.

In this section, we review and compare CDISC, the existing standard for submitting all clinical data, and FHIR, a candidate standard for submitting RWD. A summary of selected attributes of these two standards is presented in Table 1. below. In the following section, we evaluate FHIR as a possible submission standard to support future marketing applications.

Table 1. Comparison of Clinical Data Standards

	CDISC	HL7 FHIR
Attribute		
Tabulation Standard	Yes (SDTM)	Yes (FHIR)
How organized	Topic Related Domains in SDTM	Resources
Typical exchange format	SAS V5 Transport / Metadata ODM XML	JSON
Use case(s)	Submission of RCT data to regulatory authorities, including FDA and PMDA	Electronic health records, claims data
Supported FDA clinical data standard	Yes	No
Supports non-rectangular	No	Yes

structures		
Record linkage	Implicit (RELREC domain in SDTM)	Explicit (Data elements are linked via the model)
Non-standard data elements	Supplemental variables	Extensions
Analysis standard	Yes (ADaM)	No
Typical dictionaries	MedDRA, WHODrug, CDISC Controlled Terminology	SNOMED, LOINC, ICD, RxNorm
FDA review tools	Yes	No

The U.S. Core Data for Interoperability (USCDI)

The Assistant Secretary for Technology Policy / Office of the National Coordinator for Health IT (ASTP/ONC) introduced the United States Core Data for Interoperability (USCDI) as a foundational element of U.S. health data interoperability policy following enactment of the 21st Century Cures Act in 2016. The Cures Act directed the (then) ONC to establish a standardized, nationwide baseline of health data elements that certified EHR systems must be able to exchange to support patient access, care coordination, public health, and secondary uses of data. The USCDI functions as a central policy lever for advancing interoperability of healthcare data in the United States. USCDI was formally established through the ONC Cures Act Final Rule, published in 2020 and was designed to be broader, extensible, and policy-driven, with a defined process for its anticipated expansion over time.

USCDI Version 1 defined an initial set of data classes and elements such as demographics, problems, medications, laboratory results, vital signs, procedures, and clinical notes. USCDI was explicitly framed as a minimum required data set, and a transparent, stakeholder-driven process for proposing, evaluating, and adopting new data elements was established. Subsequent versions of USCDI have expanded its scope incrementally to address emerging policy priorities and interoperability gaps. With each version, ASTP/ONC has aligned USCDI more closely with modern exchange mechanisms, particularly API-based exchange using HL7 FHIR, while maintaining USCDI's role as a content standard rather than a transport or data model standard. USCDI specifies what data must be supported by the EHR, not how that data is internally represented or externally transmitted. This separation allows USCDI to define what data must be exchanged while relying on data standards and implementation guides (IGs) a to define how that data is organized and exchanged. Its evolution reflects a shift from document-centric exchange toward granular, API-driven data access and a recognition that standardized data content is a prerequisite for analytics, public health surveillance, and RWE generation.

FHIR as the Exchange Format for USCDI Data for EHRs

To support standardized exchange, ASTP regulations require certified EHRs to make USCDI data available via APIs using HL7 FHIR. FHIR defines a modular, resource-based data model and a RESTful API approach for exchanging healthcare data in a consistent, machine-readable format.

In this context, FHIR provides the exchange representation of USCDI data. Each USCDI data element implemented in the underlying EHR database must be mapped to one or more FHIR resource variables when the data is exported or transmitted. For example, USCDI clinical data such as conditions, medications, and observations may be exchanged using FHIR resources like Condition, MedicationRequest, and Observation, respectively. FHIR profiles and IGs can be used to constrain choices such as the coding system in addition to the constraints and requirements imposed by USCDI.

EHR systems rely on non-standardized internal data models, usually proprietary or semi-proprietary, optimized for transactional clinical workflows, billing, and documentation. These models often reflect legacy design decisions, relational database structures, and vendor-specific extensions developed over decades. While EHRs certified under ASTP's Health IT Certification

Program must demonstrate the ability to store and retrieve USCDI data elements, the internal representation of those elements is not standardized across vendors.

As a result, the same USCDI data element, such as “Medication,” “Problem,” or “Laboratory Result,” may be stored differently across EHR systems in terms of structure, normalization, terminology binding, and context. For example, clinical concepts may be encoded using different vocabularies internally, stored as free text with coded overlays, or represented across multiple tables or objects. These internal differences are largely invisible to end users but become very relevant when data are exchanged externally.

Critically, this means that EHR systems must transform data from their internal data models into FHIR representations when responding to API requests. This transformation layer is where semantic alignment, terminology mapping, data normalization, and contextual interpretation occur and there is no regulation of this process, only of the final FHIR format.

HL7 FHIR To CDISC Mapping Project

The HL7 FHIR to CDISC Joint Mapping Project is a collaborative initiative between HL7 and CDISC designed to bridge the gap between clinical care data (captured in EHRs via FHIR) and clinical research data (submitted to regulators via CDISC standards). Historically, clinical trial data had to be manually re-entered from a hospital's EHR into a research database (e.g., EDC). This project maps FHIR resources to CDISC variables to allow for automated eSource data retrieval. By making it easier to convert data between HL7 FHIR (commonly used in clinical systems to collect and share healthcare data) and CDISC standards, both organizations aim to reduce the barriers to using clinical information to support research. The project facilitates interoperability between HL7 FHIR and CDISC, providing a single source for mapping FHIR resources to SDTM and CDASH.

The project focuses on the most common "core" clinical domains needed for a trial. The current mapping guide covers:

- **Adverse Events (AE):** Mapping FHIR Condition and Observation resources to the AE domain.
- **Concomitant Medications (CM):** Mapping MedicationRequest and MedicationAdministration.
- **Demographics (DM):** Mapping the Patient resource.
- **Laboratory Results (LB):** Mapping Observation (specifically linking LOINC codes to CDISC lab tests).
- **Vital Signs (VS):** Mapping Observation (e.g., Blood Pressure, Heart Rate).
- **Medical History (MH) & Procedures (PR):** Standardizing how past diagnoses and surgeries flow into research tables.

Key deliverables for the project include Joint Mapping IG (https://hl7.org/fhir/uv/cdisc-mapping/) and a USCDI to SDTM Spreadsheet that maps the US Core data elements (the ones mandatory for all U.S. EHRs) directly to SDTM variables.

THE USE OF CDISC VS. FHIR

CDISC

CDISC has many attributes that make it an exceptional data standard for clinical research. FDA

first required the use of CDISC data standards for clinical and nonclinical studies that started after December 17, 2016, though the use of CDISC traces back to the early 2000s. Given that CDISC is still the current standard supported by the FDA as outlined in the FDA Data Standards Catalog, it continues to be used to submit data in support of marketing applications for medical products. Thus, FDA reviewers and analysts, SDOs, sponsors, and other industry groups are familiar with these standards. Apart from familiarity, these groups have also developed tools for the representation and analysis of CDISC data. Further, a sponsor company may have internal tools and processes established for the collection, transformation, and reporting of CDISC data. Thus, changing standards from CDISC would require a large overhaul on tools that use CDISC data for data collection, reporting, and analysis across these groups.

CDISC is a very comprehensive end-to-end standard, with data standards supporting SEND, SDTM, and ADaM data, as well as other standards such as those used for study planning (i.e., PRM) and data collection (i.e., CDASH). Unlike FHIR, CDISC also has a dedicated analysis standard to support the creation of analysis datasets and associated metadata (i.e., ADaM) and is harmonized with the SDTM. Further, there are a variety of CDISC exchange standards. For example, the Define-XML standards are used to exchange study metadata (e.g., SDTM and ADaM Define.xml files to support the submission of the study datasets). Other exchange standards include SDM-XML, ODM-XML, and CTR-XML.

Apart from data and exchange standards, additional CDISC standards include CDISC controlled terminologies and Therapeutic Area User Guides (TAUGs), which are therapeutic area-specific guidance to aid in the submission of NDA/BLA data and documentation. As of January 2024, 50 TAUGs are published on cdisc.org. In addition to TAUGs, there are numerous questionnaires, ratings, and scales (QRS) supplements published on the CDISC website which aim at the standardized representation of SDTM (and in some cases, ADaM) data for a multitude of QRS measurements. There are countless industry experts that contribute to CDISC and have provided their expertise when developing these documents. From the regulatory agency perspective, agency-authored documentation including technical specifications and the study data technical conformance guide include CDISC data. Thus, changing a data standard will result in numerous updates to the related standards and documentation for both SDOs and regulatory agencies.

CDISC data may be ‘siloed’ as submitting data specifically produced for clinical research, since it was developed for the submission of RCT data rather than for RWD. Though it has started to be used for the submission of RWD, it has not yet been optimized for the submission of RWD when RWD is submitted to support a premarket application. Further, it can be a great deal of work to convert RWD into CDISC format. Within CDISC, there are other notable considerations regarding the optimal standard for RWD. For example, the CDISC SDTM is very rigid, and it is cumbersome to add new elements not defined within the SDTM. This is difficult when considering the numerous data concepts common to EHRs and other sources of RWD that do not currently exist within the SDTMIG and makes converting RWD to submit in CDISC format difficult. Supplemental qualifier datasets are used to include additional captured data not previously defined within the CDISC SDTM. Secondly, linking is not easy within CDISC, where Related Records (RELREC) domain is needed to comprehensively link data across domains, which are both cumbersome to make as well as to analyze downstream from a regulatory perspective. Thirdly, CDISC maintains a very rectangular structure, unlike FHIR which supports non-rectangular structures.

FHIR

FHIR has many attributes which make it a viable future standard for submitting clinical data. The advantages of using FHIR for submitting all clinical data include the following:

- FHIR is optimized for RWD (and, for EHR and claims data). Over time, there is a trend for more RWD to be submitted as part of a marketing application.

- There is a trend toward collecting data for RCTs and other observational study designs in an EHR.

- Most data in EHRs are already exchanged using FHIR, and in the future, it is likely that almost all EHR data will be represented in FHIR. As a result of the 21st Century Cures Act, FHIR has already become the leading API for exchanging healthcare data. The Office of the National Coordinator for Health IT (ONC) and the Centers for Medicare & Medicaid Services (CMS) have designated FHIR as the standard required to fulfill the requirements of the 21st Century Cures Act (7). The federal rule states that certified EHRs must support the standard FHIR APIs by the end of 2022.

- FHIR provide a mechanism to link to related data within a marketing application and to external data sources that is more direct than CDISC's RELREC domain, but use of this linking has not yet been explored.

- While FHIR is not a currently supported standard by FDA for submitting clinical data, it is currently being used for several separate activities at FDA (12). For example, SPL is based on HL7's Version 3 standard and FHIR is the emerging HL7 standard that CDER is exploring to support SPL use cases. In eCTD Module 3, FHIR is also being used as the exchange standard for submitting Pharmaceutical Quality (PQ) and Chemistry & Manufacturing Controls (CMC) data (12).

While FHIR has many attractive features, there are many drawbacks to using the FHIR as a single standard to submit clinical data:

- FHIR was developed and optimized for healthcare data but it lacks many of the concepts necessary to represent clinical research data. This makes converting RCT data to the FHIR standard difficult, and in some cases, not possible. When non-RCT data is submitted in CDISC format, many non-standard variables and domains must be created, and many validation/business rules do not apply. However, FHIR is developing resources/data elements to represent research study concepts. FHIR also has several current projects that are designing FHIR profiles that can facilitate representing data for clinical research and ultimately for submission.

At this point in time, FDA and sponsors lack familiarity with FHIR and there is an expected learning curve with the use of the standard. In addition, many sponsors, CROs and other vendor have designed end-to-end processes around CDISC standards. Switching to FHIR would place a short-term burden on industry.

- We do not know what a FHIR submission package would look like (e.g., how would the data be bundled for submission, by patient or by FHIR resource? What would be the counterpart CDISC's Define.xml file to transmit metadata, a vital component needed for regulatory review?).

- FDA and sponsors would have to develop new tools or update existing tools to facilitate a FHIR submission. Furthermore, training, new processes, and substantial change management is needed to accommodate a FHIR submission. Many FDA documents and

technical specifications were developed for a CDISC submission, and these would all have to be updated to accommodate a new submission standard.

- FHIR is primarily a collection/exchange standard akin to CDISC's SDTM. Currently, FHIR does not have a model for analysis data. Until new technologies and tools are developed, analysis datasets will continue to be a key component of regulatory review.

- There is no harmonization between CDSIC and FHIR which means mapping is required. It will be challenging to create ADaM datasets when FHIR is the source format, if ADaM is used as the analysis standard and no HL7-FHIR based analysis standard is developed.

2025 FHIR FRN

On April 23, 2025, Federal Register Docket No. FDA–2025–N–0287 was posted requesting public comments "Exploration of Health Level Seven Fast Healthcare Interoperability Resources (FHIR) for Use in Study Data Created From Real-World Data Sources for Submission to the Food and Drug Administration" (13). The comment period lasted for the standard 90 days and ended on June 23, 2025. The five questions included:

1. What challenges do you see for the pharmaceutical industry regarding the current state of submitting clinical study data collected from RWD sources to FDA?

2. What opportunities and/or challenges do you see for the pharmaceutical industry on reaching a future state of clinical study data submissions collected from RWD sources using HL7 FHIR (e.g., business processes, technical considerations)?

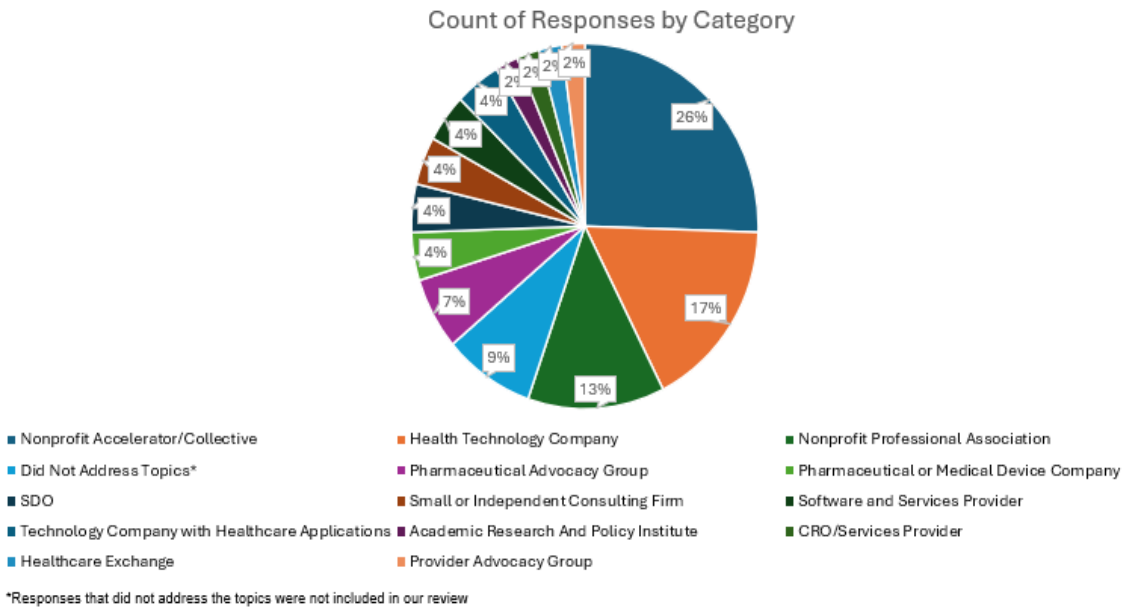
3. What are your suggestions on how, from a data standards perspective, FDA might reach a future state of clinical study data submissions collected from RWD sources that aligns with ASTP/ ONC health IT goals for HL7 FHIR-based exchange?

4. Does USCDI version 3 provide enough information for collecting RWD for research purposes? Is there information that USCDI version 3 does not sufficiently address?

5. Under TEFCA, a variety of ''Exchange Purposes'' are authorized. If ''Research'' was added as an ''Exchange Purpose,'' what role could TEFCA play with using RWD for clinical research? How could TEFCA support more efficient collection and exchange of RWD for clinical research purposes? What challenges might exist with this approach?

Of the 47 responses received, we categorized these by the type of responder in Figure 2. We evaluated responses that contained meaningful comments, which were defined as having a response provided for at least four of the five questions posed in the notice. Details of this evaluation are provided below Figure 2, separated by question.

Figure 2: FRN Responses by Category



Question 1

What challenges do you see for the pharmaceutical industry regarding the current state of submitting clinical study data collected from RWD sources to FDA?

Submitting clinical study data collected from RWD sources to the FDA presents numerous challenges to both sponsor and regulatory authorities. Currently all clinical study data must be submitted using the SDTM and ADaM CDISC standards. These CDISC standards were designed for RCTs conducted under the guidance of a protocol. Data captured during routine care is poorly aligned with CDISC standards as well as research and regulatory needs. CDISC does not contain the domains/resources and data elements to adequately represent RWD. Using CDISC, key information is not easily represented, and important variables may be missing or misclassified. Furthermore, SDTM and ADaM specify flat tabular standards which do not reflect how data from EHRs and other RWD sources are organized.

Data from EHRs and other RWD sources must be mapped and transformed to CDISC standards for submission to FDA. The data transformation process increases the risk of data loss or misinterpretation. This has the potential to undermine confidence in the data. Mapping decisions may be subjective and thus may vary widely across different groups and study processes. There is a need for a single source of “truth” or industry wide standard mappings. Additionally, extensive mappings are costly and error prone.

Interoperability of RWD with existing CDISC standards poses an additional challenge or hurdle. Not only does RWD terminology and coding differ from RCTs, but it is also inconsistent across EHRs and other RWD sources. As a result, granularity or meaning is often lost or represented inaccurately.

Transformation processes don’t always have ready-made audit trails or methods to understand how and why transformation decisions were made. Under the current state of submitting clinical study data, it is difficult to ensure end-to-end traceability (e.g., from source data to submission package), limiting the Agency’s ability to assess reliability and reproducibility.

Finally, most of the challenges of submitting RWD to FDA are not due to current submissions standards. They stem from the way RWD is collected and the business purpose of the RWD. First, a good amount of RWD is unstructured and not classified in a standard way. Second, RWD is typically incomplete, containing lots of missing data. While much of the missing data is relevant for research purposes, it does not meet business needs for patient care. In addition, RWD lacks a defined visit schedule. These factors pose significant hurdles to using RWD for clinical research, no matter what standards are used to represent the RWD.

CDISC standards also support audit trails, aggregation of data across numerous patients and data analysis. They are currently required for submission of data to FDA so that standard FDA data review tools can be leveraged. The CDISC metadata standards and associated terminology (maintained by NIH/NCI EVS) are harmonized globally, and they are freely available for researchers around the world. CDISC standards are also required by the PMDA in Japan and are endorsed by other regulators. The CDISC metadata and terminology standards are the core content standards for the current protocol standard in development by the International Council on Harmonization (ICH). The protocol representation standard and the ISO, HL7, CDISC BRIDG Model were initiated within an HL7 working group by CDISC over 20 years ago. HL7 Accelerator Vulcan is now developing the FHIR 'wrapper' for this content.

The CDISC metadata and robust terminology standards support common data elements across all clinical research studies and are present in the International Patient Summary (IPS) recommended by ISO, CEN, EU, and HL7 to transfer individual patient summaries. In addition to these foundational standards, CDISC has developed data standards and terminology relevant to over fifty therapeutic areas. These therapeutic area standards were developed in collaboration with others, including FDA reviewers (to ensure that they focus on key data elements that indicate whether or not a treatment is effective and safe), NIH, Critical Path Institute and TransCelerate through various partnerships including the Coalition for Accelerating Standards and Therapies.

Question 2

What opportunities and/or challenges do you see for the pharmaceutical industry on reaching a future state of clinical study data submissions collected from RWD sources using HL7 FHIR (e.g., business processes, technical considerations)?

Adopting HL7 FHIR as a submission standard for RWD would represent a significant shift for the pharmaceutical and regulatory ecosystem. While the long-term potential is promising, the transition introduces meaningful near-term challenges for multiple stakeholder groups, alongside opportunities to modernize evidence collection and regulatory review. Responders stated that more generally, FDA is exploring approaches to modernize submissions of clinical study data collected from RWD sources to the Agency using FHIR, while ensuring alignment with other FDA policy regarding RWD and the work of ASTP/ONC.

Question 2 further refined the scope to the question above. However, many responders envisioned future states that went well beyond this use case and even beyond current legislative mandates of FDA. This section focuses on the responses that are in scope with a brief paragraph at the end referencing how the future use of FHIR in the context of clinical studies could expand what is possible with respect to both clinical trials and post-market safety further in the future, as envisioned by a portion of the respondents. It is notable that only a few pharmaceutical companies responded to the notice and no CROs responded, though they would likely be most impacted if HL7 FHIR were to be required for submission of RWD to FDA in support of new drugs and biologics. This section also does not address issues related to the general use of RWD, such as adverse needing to reach the level where a patient seeks care from a provider to be recorded.

Q2: Mandated for Use with EHRs

HL7 FHIR was officially mandated as the data standard for use with EHRs as part of the ONC's (now ASTP) Cures Act Final Rule in the United States in the spring of 2020 with December 31, 2022 set as the compliance deadline. Many respondents cited this rule and noted that because FHIR is being widely adopted across U.S. healthcare systems, it will enable more direct integration of RWD sources at or near the point of care. While this is true for RWD derived from EHRs, it will not be true for other sources of RWD. A subset of these respondents noted EHR data is not typically stored in FHIR format and thus a conversion is needed to go from the underlying data stored in the EHR to the FHIR format. This conversion is currently unregulated While the ONC/ASTP's USCDI does provide an IG which provides guidance on what FHIR Resources and variables within those resources that must be populated, how that is done from the original EHR data is not specified. However, the hope/expectation is that conversions are similar given that they are done to exchange patient information between different healthcare systems and that the same conversion would be used to provide RWD to sponsors carrying out research.

Q2: Flexibility

The flexibility of FHIR was noted as both an advantage and disadvantage. FHIR, like CDISC allows for the creation of new variables and for RWD derived from EHR data, has fields better suited for storing relevant information, such as specific provider, location, etc. This flexibility also allows for multiple choices for mapping variables, but that can lead to the inconsistency noted above. There are two main problems that the flexibility of HL7 can yield. First, it means there are many implementations of FHIR; it varies by EHR vendors but databases that hold EHR data are customized and even health systems using the same EHR will have both a different underlying database and user interface customized to their workflow. Second, there is not currently a common understanding of data elements and their relationships (semantic alignment) across HCPs. That means different data elements could be stored in different resources within FHIR. These two areas of variability likely do not impact an individual patient where a provider is looking at the whole chart, but in cases where individual elements of data are pulled out at scale and looked at in aggregate, extra care would need to be made to ensure the right data element is being pulled from the right FHIR resource. As RWD does not have a protocol, it would require a chart review to understand if a variable was missing because a procedure didn't happen or because it was stored in a different area of the EHR than what the original ETL code was based on. Many respondents pointed out the use of FHIR would enable the use of automation and increase efficiency in data transformation and analytical pipelines. While the ability to pull FHIR data from different systems would be helpful, some efficiency will be lost in needing to ensure the EHR data to FHIR mappings are consistent. One responder noted that the *"technical burden of ETL and data wrangling is perceived as remaining at current levels, or possibly increasing, with the adoption of FHIR and the increased prevalence of RWD submissions."* It may warrant some investigation, especially for large data aggregators, to understand if and how this issue could be best addressed.

Q2: Coding Variability

A third area of variability arises in what one respondent referred to as "depth" but could also be referred to as specificity and is related more to the use of RWD than to FHIR specifically. A responder shared that "*The depth of FHIR implementation varies across hospital sites. [S]ome hospital sites may map a procedure to a specific code (e.g. "spirometry") and others map to the least specific code of "diagnostic procedure". Both are compliant, but one is usable for research and the other is not."*

One responder noted that *"[e]ven when data are structured, inconsistent use of vocabularies (SNOMED vs ICD-10, local codes, uncoded strings) leads to ambiguity and loss of clinical meaning [creating] unknown unknowns for FDA reviewers and erod[ing] confidence in analyses."*

Another responder noted a related but different issue at the critical point at which healthcare data (RWD) are accessed for secondary use (e.g., research, safety surveillance, public health). This is where the terminologies must be clearly interpreted and the metadata unambiguous; however, it is unclear if this is possible given the circumstances of collection during routine care. For example, a clinician who is pressed for time and must choose LOINC codes which are "pre-coordinated" must choose between the more than 60 codes for glucose, and/or a SNOMED code which has a multitude of codes (accessed via a browser) for a given data element. SNOMED CT contains more codes compared to MedDRA. A CRA is trained to review the data and has the ability to reach out to an investigator to get more details on a particular code choice, not to mention that investigators are trained by sponsors to ensure consistent data.

The FDA currently does not have a way of being able to know the level of specificity of contributing sources of RWD, although it is something the sponsor could (and should) investigate prior to using the data. Whether the data is fit for purpose would depend on the efficacy and/or safety endpoint. For overall survival rate, this depth/specificity may not matter but, using the above example of spirometry vs diagnostic procedure, for a drug that may have adverse events on the lungs, it may be something FDA would want to know about. One responder noted that *"unconstrained FHIR will recreate today's bespoke mapping chaos".*

Q2: Provenance

FHIR's provenance resource and other extensions allow for structured metadata to trace data lineage, source system, and data creation method. In addition, the USCDI's IG requires the use of FHIR's provenance resource.

The current FHIR standard is limited in its ability to provide clinical context and workflow metadata, such as context regarding data capture (e.g., order provenance) and clinical workflow steps. FHIR resources currently don't contain detailed metadata about how and why data were recorded, which could impact bias detection or quality grading of the data. Note that CDISC also currently has a traceability working group investigating how to provide RWD traceability using a define.xml extension file.

Q2: Burden to Stakeholders/Industry Readiness

Several comments focused on stakeholder readiness noting that clinicians, sites, CROs, and regulators may lack familiarity with FHIR and FHIR-based workflows and require training. Assuming the traditional submission process, which this FRN does, only CROs, sponsors, and regulators as well as those handling the data for analysis and FDA submission should be impacted by the data collection. If, in the future, data is collected through EHRs and not through CRFs, that change would impact clinicians and sites. However, people who do handle the data, analyze the data, and format/process the data for FDA submission would need training on how to interpret and analyze data in the new model. The learning curve for FHIR has been shown to be quite steep (14).

The responders also noted the effort and investment needed to build and maintain technical infrastructure required for FHIR. Again, assuming normal workflow, a FHIR server would not necessarily be required to collect FHIR data but could be an option. The effort would likely depend on how the data was received. If buying data from an aggregator who does not use FHIR format, sponsors could require the aggregator to transform the data before delivery. Insight from the transition to CDISC standards might be helpful here, although hopefully some lessons were learned during that process and transition to a new standard will not be as difficult as the initial transition to CDISC. In addition, many sponsors have used RWD for trial feasibility studies and perhaps have some infrastructure already set up that could also be used. Unfortunately, only two pharmaceutical companies and no CRO responded to this notice and no specific estimate of the technical burden can be intuited from the responses given. CDISC SDTM is designed to be

stored in a relational database while FHIR, with its more flexible structure, is not. Storage of the FHIR data may present a new technical challenge to pharma but an insurmountable one.

Ongoing adaptation to evolving FHIR guidelines was cited as a challenge which might then require business processes be reengineered. Similarly to CDISC, FHIR uses the concept of IGs to specify specific resources and variables within those resources that must be used, as well controlled vocabularies. A change or update to the FHIR IG, while a meaningful change, should not require a complete rework of the technical infrastructure just like an update the CDISC IG should also not require a rework of technical infrastructure.

One responder noted that the amount of cost and time savings from using HL7 FHIR over CDISC hinges on how widespread adoption is and how efficiently infrastructure can be developed to seamlessly gather and organize RWD across healthcare data systems. The entire industry would need to adopt new programs, platforms, and pipelines to maximize savings, but this comes with significant up-front costs. Many legacy systems are not designed to handle structured FHIR data and would require significant upgrades.

Q2: FDA Readiness

Responders noted that there are no tools to handle the analysis of FHIR data but considered this as an opportunity for FDA and industry to work together on tools for the validation or analysis of RWE submitted in FHIR format. A validation tool would need to be built but is not something that is prohibitive. In the early days of CDISC, some may remember that validation handled by proprietary tools before OpenCDISC/Pinnacle 21 entered the scene. FDA leverages review tools (e.g., MAED) and data scientists who work with medical reviewers to create safety analyses for them. Efficacy analyses are handled by biostatisticians. Both groups are well versed in data and data transformations and would likely not be greatly impacted by this change, although this should be confirmed.

Q2: Analysis Standard

Several responders noted that FHIR is an ingestion layer, akin to CDISC SDTM and not a final analytic format like ADaM. FHIR does not have an analysis standard, though it has been suggested that one could be created using the observation resource. One responder noted that there is potential to guide a process/standard to map HL7 FHIR directly to ADaM but as of this paper, no pilot for this has been done.

Some may remember that the ADaM standard followed behind the SDTM standard. Analysis was originally done from the SDTM standard and later evolved to use the ADaM standard. An initial approach may be to pilot the use of ADaM with FHIR data to see if another analysis needs to be created. This could also solve some issues in merging CDISC and FHIR datasets although the merger would only be able to be done at the analysis level.

Q2: Semantic Interoperability

Given the diversity for data sources and even the diversity among a common source (e.g., data may be collected differently across EHRs and even across different implementations of the same EHR) it is difficult for researchers to use RWD and a number of models have been created and are currently in active use. These models include PCORNet, OMOP, i2b2 and Sentinel which we covered in our earlier papers /presentations with the expectation of i2b2.

One responder emphatically addressed semantic interoperability and felt that *“unless the semantics are adequately addressed, interoperability will not work, or at least it will not yield trustworthy results.”* The recommendation was then *to “leveraging the synergy between two standards”* using HL7 FHIR as a transport standard carrying CDISC content (i.e. data in the

format of CDISC metadata and terminology standards). FDA has as the author noted, a project dedicated to creating Interoperability of Common Data Models and Data Standards" (i.e. CDMH Phase 3).

Challenges in this area related to the importance of semantics and metadata/terminology standards and the difficulty of development. In the opinion of the responder, rather than start from scratch, those that are already available through NCI EVS, CFAST and CDISC should be leveraged. There was also concern that attempts to develop new FHIR resources for therapeutic areas will create redundancy and could 'break' the existing standards and thus compromise interoperability

Q2: Standard Design Aims

FHIR was developed to exchange data by patient, it was not originally suitable for aggregating and transporting data from multiple patients and was not designed to facilitate the tabulation and statistical analysis which are critical to determining whether research results are statistically significant and thus trustworthy. CDISC was developed using specific domains relevant to clinical trials to support the entire clinical research process, from protocol and data collection through aggregation, tabulation, and analysis.

Q2: Future-Future State

The FRN was very clear about the context of the questions we address below focusing on submitting clinical study data collected from RWD sources to FDA. Despite the clarity of the questions, many responders covered use cases that were not included in this FRN. This primarily took the form of assuming the FDA is releasing this FRN with the intention to collect real time data from EHRs, some specifically noted the use case for post-market surveillance and others did not mention this use case, so it assumed they intended this use for NDA/BLA approval since that is what the question specified. One responder did directly comment on the impact it could have on new drug trials, stating that promise of near real-time access to clinical data, enabling adaptive trial designs and earlier safety/efficacy insights. Clinical research is currently done by the sponsor and reviewed by FDA and while specific cases, like the recent pandemic, might make it expedient for FDA to harness RWD for this use, it is not currently standard for them to do so until a drug is marketed. That said, it is perhaps with that future in mind that FDA is looking towards using the FHIR standard with submitted RWD as a path towards the ability to perform real-time analysis and is choosing to look at FHIR over other standards (like OMOP or Sentinel) that would require a data transformation before use with post-market surveillance. However, this future state of real time clinical data is not a subject the FDA requested to be addressed in the FRN.

Q2: Summary

Responses to FDA's exploration of HL7 FHIR for submission of clinical study data derived from RWD reflect both cautious optimism and significant concern. Responders broadly agree that FHIR's mandated adoption for EHR interoperability positions it as a logical candidate for sourcing RWD, particularly data originating from U.S. health systems. Its modern API-based architecture and growing ecosystem create opportunities to streamline data access and potentially modernize regulatory submissions over the long term.

At the same time, responders consistently emphasized that FHIR, as currently implemented, is not a turnkey solution for regulatory-grade clinical study submissions. Variability in EHR implementations, unregulated EHR-to-FHIR transformations, inconsistent vocabulary usage, and differences in data specificity introduce risks to data quality, semantic consistency, and fitness-for-purpose. These issues are amplified when data are extracted at scale for secondary use, where missing context, ambiguous provenance, and heterogeneous coding practices may impact analysis results.

Several responders stressed that FHIR functions primarily as an ingestion or exchange layer, comparable to CDISC SDTM, and not as an analysis-ready standard. The absence of a defined analytic model analogous to ADaM, combined with limited tooling for validation and review of FHIR-based submissions, underscores that substantial infrastructure, standards development, and workforce training would be required across industry and FDA. While FHIR offers mechanisms for provenance and metadata, current implementations may not capture sufficient clinical context or workflow information to fully support regulatory decision-making for new medical products.

Overall, responders viewed the use of FHIR for RWD submissions as feasible only if accompanied by stronger constraints, clearer semantic alignment, and deliberate coordination with existing clinical research standards rather than as a wholesale replacement for them. Many cautioned that unconstrained adoption could recreate existing data wrangling challenges under a new technical paradigm.

Q2: Recommendations

Based on the responses reviewed, several recommendations emerge to support a responsible and effective path forward for the use of HL7 FHIR in submissions of clinical study data derived from RWD:

1. Constrain FHIR Implementations for Regulatory Use: FDA will need to define a regulatory-specific FHIR IG for RWD submissions. This guide should reduce variability by constraining resource usage, required elements, and vocabulary bindings relevant to regulatory endpoints, while remaining aligned with USCDI for common data elements but will need to address submission of resources beyond what USCDI covers.

2. Address Semantic Interoperability Explicitly: Semantic alignment should be treated as a first-order requirement. Rather than developing new therapeutic-area FHIR resources, respondents recommend leveraging existing CDISC metadata, terminology, and NCI EVS assets. One respondent noted potentially using FHIR as a transport mechanism for CDISC SDTM-aligned content. Ongoing (or previous) FDA efforts such as the Common Data Model Harmonization initiative could be used to bridge FHIR with established research standards.

3. Clarify Expectations for EHR-to-FHIR Transformations: Given that EHR data are rarely stored natively in FHIR format, greater transparency and consistency in transformation logic is critical. FDA should consider whether sponsors (and by extension 3rd party aggregate data providers) need to understand and document how source EHR data are mapped to FHIR, assess variability across contributing sites, and evaluate fitness-for-purpose relative to specific regulatory questions.

4. Pilot Analysis Pathways Rather Than Assume Net Efficiency Gains: Before assuming cost or efficiency benefits, FDA and industry should pilot limited use cases that explore what a FHIR submission would look like (e.g., how bundling would be done) and the impact to the current FDA workflow. This would include validating the FHIR submission and determining how to perform common safety analyses. FDA biostatisticians should also have the opportunity to assess the data for impact to their workflow. This could include a test mapping of FHIR data into the CDISC analysis standard ADaM. These pilots could help determine whether a new analysis standard is necessary or whether existing approaches can be adapted without increasing analytic burden.

5. Invest in Tooling, Training, and Phased Adoption: Respondents highlighted the need for validation tools, reviewer-facing analysis utilities, and workforce training. FDA should the experience of the transition to CDISC to anticipate the support is needed for both internal and external stakeholders. A phased approach, with voluntary pilots and clear feedback loops,

would allow FDA and industry to mature capabilities before the submission of RWE grows more common.

6. Maintain Clear Separation Between Near-Term and Future-State Use Cases: While some respondents envisioned real-time data access and post-market surveillance applications, these future-state scenarios extend beyond the scope of the FRN. FDA should clearly distinguish near-term expectations for RWD submissions supporting regulatory decisions from longer-term ambitions, to avoid misalignment of investments and assumptions.

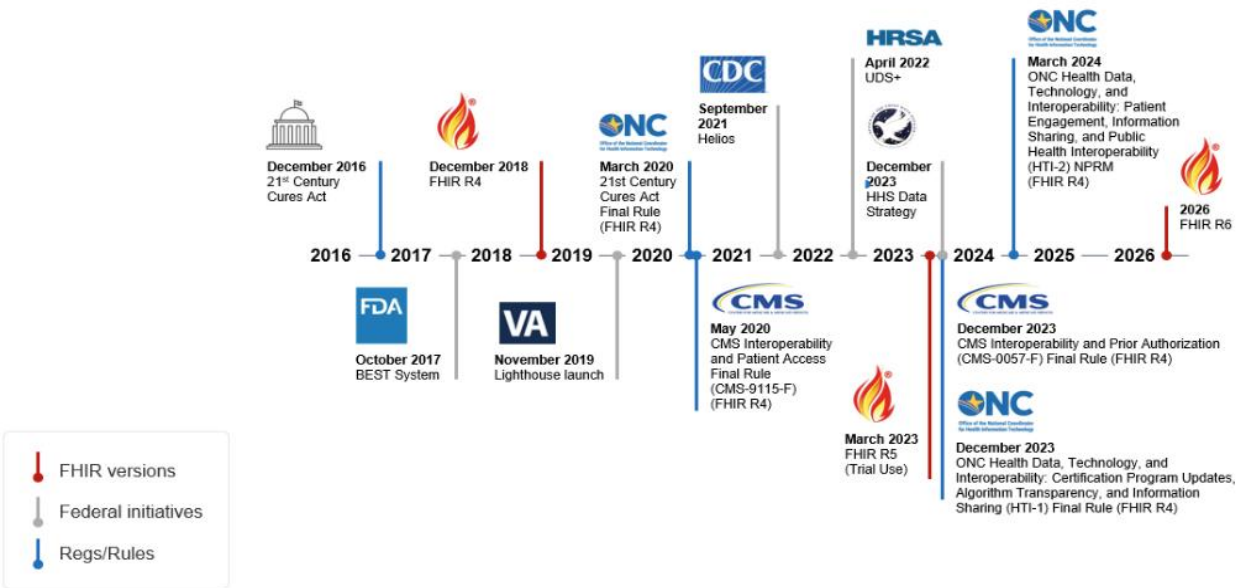
Taken together, the responses suggest that most stakeholders are optimistic about the use of HL7 FHIR for RWD-based regulatory submissions, but only if adopted as part of a carefully governed, semantically grounded ecosystem that builds on, rather than replaces, decades of clinical research standards development.

Question 3

What are your suggestions on how, from a data standards perspective, FDA might reach a future state of clinical study data submissions collected from RWD sources that aligns with ASTP/ONC health IT goals for HL7 FHIR-based exchange?

The 2024 Draft Federal FHIR Action Plan, released by the Assistant Secretary for Technology Policy/Office of the National Coordinator for Health Information Technology (ASTP/ONC), guides federal agency investment and adoption of HL7 FHIR, aiming for progress by 2026. The Action Plan names six HL7 FHIR components that are specifications they will use to lay the foundation for the next generation of health care interoperability. One of the six components are “Research Components intended to drive toward a fully digital health system that uses FHIR for research activities.” (15). Right now, two are named in the plan: the HL7 FHIR mCODE (minimal Common Oncology Data Elements) Implementation Guide (mCODE IG) and the Standard / Specification: HL7 FHIR Genomics Reporting Implementation Guide (Genomics Reporting IG).

Figure 3. Draft Federal FHIR Action Plan Graphic, showing interoperability mandates and FHIR adoption progress



As the FDA continues to explore the use of RWD to support regulatory decision-making, alignment with ASTP/ONC health IT goals and other HHS agencies, including HL7 FHIR–based exchange, provides an opportunity to streamline clinical study data submissions and align them with new health related data initiatives from other agencies.

Q3: Harmonized Standards and Alignment with ONC Initiatives

FDA should encourage SDOs, including CDISC and HL7, to establish comprehensive and harmonized standards for RWD and RWE. Building on complementary initiatives already underway in clinical research such as HL7 Vulcan projects can reduce duplication and accelerate convergence between healthcare and research data models. Anchoring FDA RWD submission guidance to USCDI, while remaining aligned with ONC-driven FHIR efforts, would further ensure consistency with national health IT policy and promote interoperability across EHR systems and research workflows.

A practical step toward this alignment would be the development of a “RWD-Core” FHIR IG that defines a core set of required data elements and controlled terminologies for regulatory-grade RWD submissions. Such an IG would maximize the use of FHIR as a submission standard while allowing extensions for therapy-specific or use-case-specific needs.

Q3: Provenance, Traceability, and Data Quality

Robust provenance and traceability are foundational to provide confidence in submitted RWD. FDA should clarify provenance requirements more thoroughly, including specifying minimum access, auditability, and lineage standards across the full transformation pathway from source EHR data to the submitted RWD. Clear guidance for documenting provenance within FHIR submissions, potentially through standardized metadata fields or FHIR extensions, would help address the complexity and heterogeneity of RWD sources.

In parallel, FDA should adopt a fit-for-purpose approach to data quality, provenance, and required standards, adjusted for the intended regulatory use. This approach, consistent with recent FDA and TransCelerate discussions on pragmatic elements at scale, would reduce unnecessary burden while ensuring appropriate rigor when data are used for high-impact regulatory decisions. Requiring documentation of data capture, validation, and transformation methods used across the data lifecycle, as well as requesting ASTP/ONC recommend EHR vendors embed quality assurance features at the source, would further strengthen data reliability.

Q3: Bridging FHIR and CDISC Through Guidance, Tooling, and Pilots

Given FDA’s continued reliance on CDISC standards such as SDTM, ADaM, and Define-XML, there is a clear need for definitive guidance and tooling to map FHIR resources to research-ready formats. FDA should convene a multi-stakeholder effort to endorse an existing framework (or develop a new one) for FHIR-to-CDISC transformations, including clear validation pathways that preserve provenance and traceability.

Conducting targeted pilots in partnership with FDA, supported by focused, multidisciplinary think tanks with pharmaceutical representation, would allow these mappings and standards to be tested in real-world scenarios. Grouping and analyzing large volumes of protocols, potentially aided by AI, could help identify the most common data elements required for specific use cases, informing both FHIR, USCDI(+), and upstream EHR data capture priorities.

Q3: Flexible Adoption Models and Enablement

Rather than forcing full-FHIR submissions, FDA may consider adopting a hybrid model that balances modern interoperability standards with existing regulatory processes and that leverages

FHIR's modern capabilities. Maintaining flexibility by supporting multiple data standards that are suited to different data sources and purposes will be essential during the transition period. Alternatively, FDA may explicitly recognize HL7 FHIR as an acceptable standard for RWD submissions to encourage greater interoperability between health IT systems.

To accelerate adoption, FDA could encourage the creation of standardized, open-source synthetic test datasets for RWD submissions, enabling sponsors and technology vendors to validate implementations early.

Q3: Collaboration, Certification, and Regulatory Innovation

Close collaboration between FDA and ASTP/ONC will be critical to integrating FDA data quality and GCP-aligned requirements into broader health IT goals. FDA could further strengthen the ecosystem by developing certification criteria for technologies that bridge FHIR-based health data and research-ready formats, and by establishing a regulatory sandbox to test innovative approaches, including privacy-preserving technologies. Finally, FDA's continued active participation in and sponsorship of community efforts such as FHIR Connectathons and early adopter programs would signal commitment, reduce industry risk, and accelerate the pharmaceutical sector's transition toward FHIR-based RWD submissions.

Q3: Conclusion

By promoting harmonized standards, clarifying provenance and data quality expectations, enabling robust FHIR-to-CDISC mappings, and adopting a flexible, fit-for-purpose approach, FDA can align RWD submissions with ASTP/ONC health IT goals while preserving regulatory rigor. Strategic collaboration, piloting, and ecosystem investment will be essential to realizing a future state in which FHIR-based exchange supports scalable, high-quality, and interoperable regulatory use of RWD.

Question 4

Does USCDI version 3 provide enough information for collecting RWD for research purposes? Is there information that USCDI version 3 does not sufficiently address?

Across all respondents who addressed this question, there is broad consensus that USCDI version 3 is insufficient for collecting RWD for research purposes. While some felt USCDI v3 is an important foundational step for establishing core data elements, it is widely characterized as a baseline rather than a ceiling and remains too high-level to support advanced research needs.

Respondents consistently note that USCDI v3 lacks sufficient detail, granularity, and longitudinal depth required to assess fitness for purpose and to generate regulatory-grade RWD. Critical gaps include missing data elements for disease-specific contexts, such as oncology attributes including tumor size, grade, histology, staging, line of therapy, and outcome measures. Similarly, therapy details, outcome data, and granular medication administration and pharmacovigilance information are not adequately represented and insufficient for research. One positive response noted that USCDI does mandate that FHIR Provenance Resource be included in FHIR Bundles facilitating traceability.

In addition, USCDI v3 lacks detailed data elements like unique device and test kit identifiers for the laboratory domain, which are essential for validating, integrating, and deduplicating laboratory results.

Even when elements were defined, respondents noted concerns over unclear or non-intuitive data classes, such as "Goals" and "Problems," which may reflect narrow clinical workflows and fail to capture clinically meaningful attributes like condition severity. Respondents also note that

USCDI is not aligned with international standards and practices, limiting its usefulness for global research programs.

There was concern about USCDI's focus on structured EHR data constraining advanced analytics and AI applications that depend on richer, more contextual data. In addition, that RWD can be from any source, would these elements be meaningful in other contexts, e.g., for RWD from claims or registries. One responder noted that USCDI only covers a subset of Electronic Health Information (EHI) and only defines the core concepts and their required vocabulary.

Temporal information is another major limitation. USCDI v3 does not consistently capture actual dates, relative timing, or temporal relationships necessary for longitudinal analyses, event sequencing, and outcome attribution. In addition, structured information on patient enrollment, follow-up, and observability is lacking, further constraining longitudinal research.

Several responses emphasize that USCDI v3 lacks a terminology and code list layer. While technically true, this is by design USCDI is a content standard (and is actually technology agnostic) and it does reference which vocabulary systems apply (e.g., SNOMED CT, LOINC, RxNorm, UCUM), but it does not provide value sets. The US Core IG is a technical implementation specification built on HL7 FHIR and defines FHIR profiles for the USCDI elements and this IG binds specific value sets and code systems to the variables. Like CDISC, these can be either recommended or required.

Responders noted that USCDI is not consistently implemented across sites, with some sites choosing to capture more granular or specialized data than others. The different "flavors of implementation" can lead to inconsistencies when the data is aggregated across healthcare organizations and laboratories. This implementation variability further weakens USCDI's utility for research. For pharmaceutical sponsors, interaction with USCDI is indirect, mediated through the clinical sites they engage with, and it may not be apparent how the specific implementation of USCDI impacts what data their sites can or cannot provide.

One responder expressed concern that an increase in USCDI elements for the research use case would cause significant administrative burdens and unfunded mandates related to data collection, reporting, and compliance. Additional requirements, especially those not aligned with existing workflows or requiring costly technology upgrades, could worsen these burdens and divert essential resources away from direct patient care.

Q4: Summary

In summary, stakeholders were almost unanimous in agreeing that USCDI version 3 does not meet the needs of clinical research and regulatory RWD collection. They were more mixed whether they thought USCDI could be updated to handle these needs with some giving recommendations for specific updates and some noting that this really a moving target and USCDI will likely be missing elements given the inherent complexity and variability of research use cases. Some responders see its greatest value lies in capturing commonly used core data such as demographics, diagnoses, problems, and routine laboratory results.

Efforts to substantially expand USCDI for research risk increasing administrative burden, compliance costs, and workflow disruption, potentially detracting from patient care. Although some mentioned developing a research-oriented extension, such as a 'USCDI-R', the data then would likely not be RWD which is defined as data captured in the context of routine care. While future technology may reduce the impact of data collection on clinical practice, current systems have not yet reached that point.

Q4: Recommendations

There is strong support for continuing the expansion of USCDI to focus on core elements in addition to keeping the FHIR provenance resource as part of the USCDI profile. This approach is similar to how CDISC evolved, focusing first on standardizing core elements into specific domains (e.g. demographics, exposure, adverse events, disposition) and only later widening their focus to include therapeutic areas.

Several responses suggested that FDA could anchor RWD submission guidance to USCDI v3 while defining a more explicit “RWD-Core” FHIR IG for regulatory use cases. FDA’s RWD project, which ended in Sept 2025, had already begun the work of developing draft profiles and IGs for a FHIR based submission and the FDA is aware that a FHIR based submissions would require these profile(s) and accompanying IG(s) be developed before using FHIR has a standard for submission. Another responder suggested leveraging the USCDI profile and IG as a starting point and pointed to the USCDI’s mandate for including Provenance resources in FHIR Bundles which could help address gaps related to traceability and data lineage, although the current FDA Commissioner is considering removing the traceability requirement from RWD submissions. Any updates to USCDI should avoid negatively impacting the direct patient care from which RWD is drawn.

RECOMMENDATIONS

The increasing reliance on RWD to support regulatory decision-making presents both an opportunity and a risk. While emerging interoperability standards such as HL7 FHIR offer a promising pathway to modernize the submission of RWD, the responses to FDA’s 2025 FRN make clear that a careful, staged, and harmonized approach is essential. Based on the evaluation of stakeholder feedback and the analysis presented in this paper, the following recommendations are proposed to support an effective and sustainable future state for submitting RWD in support of NDAs and BLAs.

Adopt a Constrained, Fit-for-Purpose Use of FHIR for Regulatory Submissions: FHIR should not be adopted as an unconstrained or wholesale replacement for existing regulatory data standards. Instead, FDA should define a regulatory-specific FHIR implementation framework for RWD submissions that constrains resource usage, required elements, and vocabulary bindings based on regulatory needs. This framework should align with USCDI for foundational data elements while extending beyond USCDI to support regulatory endpoints, longitudinal analyses, and safety evaluations. A clearly defined “RWD-Core” FHIR IG would reduce variability, improve consistency across submissions, and mitigate the risk of recreating current data wrangling challenges under a new technical paradigm.

Preserve and Leverage Established Clinical Research Semantics: Semantic interoperability should be treated as a first-order requirement. Rather than developing parallel or duplicative therapeutic-area constructs within FHIR, FDA and standards development organizations should leverage existing CDISC metadata and controlled terminology. Harmonization efforts, such as the HL7 FHIR–to–CDISC mapping initiative and FDA’s Common Data Model Harmonization activities, should be expanded and operationalized. Using FHIR as an exchange or transport layer for semantically grounded research content offers a pragmatic path forward that builds on decades of clinical research standards development.

Clarify Expectations for Data Transformation, Provenance, and Traceability: Given that most EHR data are not stored natively in FHIR format, the unregulated transformation from source systems to FHIR representations represents a critical risk to data reliability. FDA should clarify expectations regarding documentation of EHR-to-FHIR transformations, including transparency into mapping logic, variability across contributing sources, and assessment of fitness for purpose relative to specific regulatory questions. Provenance and traceability requirements should span the full data lifecycle from source capture through transformation and submission and be explicitly articulated for FHIR-based submissions.

Pilot Submission and Analysis Pathways Before Mandating Broad Adoption: Assumptions regarding efficiency gains, cost savings, or analytic readiness should be validated through targeted pilot programs. FDA and industry should collaboratively pilot limited RWD submission use cases to evaluate how FHIR-based data would be bundled, validated, reviewed, and analyzed within existing regulatory workflows. These pilots should include exploration of mapping FHIR-derived data to CDISC ADaM to assess whether existing analysis standards can be retained or adapted, or whether additional analytic models are necessary.

A pilot comparing the use of a single data standard versus hybrid approach should be done. The authors of this paper pose the question of whether CDISC or FHIR can adequately represent all data used for pre-market evidence of medical product efficacy and safety should be considered along with the question of interoperability considering the increasing diversity of study designs as innovations in clinical research continue to be made.

Using a single standard option encompasses choosing either the CDISC SDTM or HL7 FHIR standard to represent all study designs for a marking application. In this scenario, the collected data (or tabulation data) is submitted in the tabulation standard chosen by FDA, with input from industry. However, the pros and cons outlined in the preceding section must be considered for each standard, as these will be relevant for NDA/BLA data submitted under the selected standard.

In the hybrid approach, RCT data are submitted using the SDTM standard, study data derived from RWD submitted using the FHIR standard. A benefit of using a hybrid approach is allowing data to be submitted in the standard for which it is optimized (e.g., CDISC for RCT data, FHIR for RWD), but a drawback is the challenge of pooling this data due to the lack of harmonization between CDISC SDTM and HL7 FHIR. Having multiple standards requires two sets of tools and processes and can make the analysis of RCT data and RWD difficult if data are not harmonized to facilitate aggregation. Currently, with no analysis standard for FHIR, the CDISC ADaM standard would be the logical choice for submitting analysis data. However, CDISC SDTM was created prior to ADaM, and analysis can be done on FHIR data as it was once done on SDTM data with the FHIR analysis standard development following. However, analyzing FHIR data in aggregate as well as creating ADaM datasets from FHIR as the source format should be piloted to understand the feasibility of both efforts.

Involvement of all stakeholders, including FDA reviewers, biostatisticians and industry will be essential to ensure that new submission formats do not inadvertently increase data submission preparation or review burden.

Invest in Tooling, Training, and Phased Enablement: Meaningful adoption of FHIR for regulatory submissions will require investment in validation tools, reviewer-facing analysis utilities, and workforce training across sponsors, CROs, and regulatory agencies. FDA should draw on lessons learned from the transition to CDISC standards by supporting phased adoption, voluntary pilots, and clear feedback loops. Open-source tools, synthetic test datasets, and shared reference implementations could lower barriers to entry and promote consistent implementation across the ecosystem.

Maintain Clear Separation Between Near-Term Regulatory Use and Future-State Aspirations: While some stakeholders envision future scenarios involving near–real-time data access, adaptive trials, or expanded post-market surveillance, these use cases extend beyond the scope of current regulatory submissions. FDA should clearly distinguish near-term expectations for RWD submissions supporting NDA and BLA decisions from longer-term ambitions. This separation will help align investments, manage stakeholder expectations, and ensure that immediate efforts remain focused on improving the quality, interpretability, and reviewability of submitted evidence.

In summary, stakeholder responses reflect cautious optimism about the role of HL7 FHIR in enabling regulatory use of RWD, provided it is implemented within a carefully governed, semantically grounded, and harmonized framework. A hybrid approach (one that modernizes data exchange while preserving proven clinical research standards) offers the most credible path to ensuring that RWD can reliably support timely regulatory decisions and, ultimately, improve patient access to safe and effective medical products.

CONCLUSION

The increasing use of RWD to support regulatory decision-making reflects a broader shift in how clinical evidence is generated and evaluated. As demonstrated by stakeholder responses to the FDA's 2025 FRN, existing submission standards are not fully optimized for the scale, structure, and variability inherent in RWD. HL7 FHIR offers a compelling foundation for modernizing data exchange from healthcare systems, particularly for data originating in EHRs, but it is not yet a complete or analysis-ready solution for regulatory submissions. Variability in implementations, inconsistent terminology usage, and limited analytic tooling present substantial challenges that must be addressed before widespread adoption can be realized.

The findings of this review suggest that a hybrid and evolutionary approach, rather than a wholesale replacement of CDISC standards, is the most pragmatic path forward. Leveraging FHIR for data ingestion and exchange while preserving established clinical research semantics, metadata, and analysis models can help balance innovation with regulatory rigor. Clear guidance on provenance, traceability, and transformation processes will be essential to maintaining confidence in RWD used for high-impact regulatory decisions. Incremental pilots, constrained IGs, and continued collaboration among FDA, standards development organizations, and industry stakeholders will be critical to refining this approach.

Ultimately, the successful integration of RWD into regulatory submissions depends not only on technical standards but on governance, transparency, and shared understanding of fitness for purpose. By aligning health IT interoperability initiatives with longstanding clinical research standards, the regulatory ecosystem can evolve to support more efficient, reliable, and scalable evidence generation, ensuring that safe and effective medical products reach patients as expeditiously as possible.

REFERENCES

1. Abolafia, J, Ferko, S, & Holt, I. (2022). "Submission Standards for RWD: The Good, the Bad and the Ugly". Paper presented at the PHUSE Annual Conference 2022, Belfast, United Kingdom. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2022/Connect/EU/Belfast/PRE_RE09.pdf
2. Ferko, S., Holt, I., & Abolafia, J., (2023). "Challenges and Considerations for Submitting Real World Data". Paper presented at the PHUSE US Annual Conference 2023, Orlando, FL. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2023/Connect/US/Florida/PRE_RE05.pdf
3. Abolafia, J, Ferko, S, & Holt, I. (2023). "Submission Standards for Real World Data: Gaps, Limitations and Recommendations". Paper presented at the PHUSE Annual Conference 2023, Birmingham, United Kingdom. https://phuse.s3.eu-central-1.amazonaws.com/Archive/2023/Connect/EU/Birmingham/PAP_RE03.pdf
4. Aetion. "Aetion_eBook_2021_The role of RWE in FDA Approvals". https://aetion.com/evidence-hub/welcome-to-the-evidence-hub/
5. Purpura CA, Garry EM, Honig N, Case A, Rassen JA. The Role of Real-World Evidence in FDA-Approved New Drug and Biologics License Applications. Clin Pharmacol Ther. 2022 Jan;111(1):135-144.
6. FDA Guidance (2018). "Framework for FDA's Real-World Evidence Program". December 2018. https://www.fda.gov/media/120060/download
7. 21st Century Cures Act. (2016). https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf

8. FDA Guidance (2021). "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision Making for Drug and Biological Products". September 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/real-world-data-assessing-electronic-health-records-and-medical-claims-data-support-regulatory
9. FDA Guidance. (2021). "Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry". October 2021. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-drug-and-biological-product-submissions-containing-real-world-data
10. FDA Guidance. (2023). “Data Standards Catalog”. April 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-catalog
11. FDA Eliminates Major Barrier to Using Real-World Evidence in Drug and Device Application Reviews, https://www.fda.gov/news-events/press-announcements/fda-eliminates-major-barrier-using-real-world-evidence-drug-and-device-application-reviews#:~:text=The%20FDA%20is%20responding%20to,traditional%20clinical%20trials%20cannot%20capture. *(accessed 2/8/2026)*
12. Galvez, J., (2023). “Reimaging Regulatory Data Submissions – Modernizing, partnering and exploring how regulatory data could be shared with HAs””. Paper presented at the PHUSE US Annual Conference 2023, Orlando, FL.
13. U.S. Food and Drug Administration. (2025, April 23). *Exploration of Health Level Seven Fast Healthcare Interoperability Resources for use in study data created from real-world data sources for submission to the Food and Drug Administration; Establishment of a public docket; Request for comments* (Notice No. FDA-2025-N-0287), 90 Fed. Reg. 17067–17069. https://www.federalregister.gov/documents/2025/04/23/2025-06967/exploration-of-health-level-seven-fast-healthcare-interoperability-resources-for-use-in-study-data
14. The Association of State and Territorial Health Officials (ASTHO), “Accelerating Data Exchange in Public Health, Insights from Public Health FHIR® Pilots”, July 2025.
15. Assistant Secretary for Technology Policy, The Draft Federal FHIR Action Plan, https://isp.healthit.gov/about-fhir-action-plan (*accessed 2/6/2026*)

CONTACT INFORMATION

Ingeborg Holt

Orizaba Solutions LLC

202-528-0020

iholt@orizabasolutions.com

www.orizabasolutions.com