PHUSE EU Connect 2024

Paper RE03

Considerations for the Submission of RWD using CDISC with Insights from HL7 FHIR and OMOP

Jeff Abolafia, P21, Radnor, PA, USA
Sarah Ferko, IBM Consulting, USA
Ingeborg Holt, Orizaba Solutions, USA

ABSTRACT

There is a rapid increase in the use of RWD to support marketing applications. However, submitting RWD may pose numerous challenges for regulatory reviewers in their analysis of the data. In our four previous papers, we outline these challenges and present several high-level solutions. This current paper builds on our previous work by taking a deep dive into an example CDISC SDTM submission and highlighting updates needed to the SDTM to better accommodate RWD. Specifically, we examine challenges and considerations for the Exposure domain within this paper. Considering how other models such as HL7 FHIR, OMOP, and PCORnet handle certain elements, we propose new solutions that are needed to support RWD. These ideas are intended to help sponsors submit RWD and to start the conversation with SDOs such as CDISC around the updates needed for more efficient regulatory review.

INTRODUCTION

Over the past several years, there has been a sharp increase in the number of New Drug Applications (NDAs) and Biologics License Applications (BLAs) that include Real World Data (RWD) and data from other observational studies. Approximately twenty years ago, when CDISC standards were initially developed to support marketing applications, NDAs and BLAs were primarily comprised of a collection of randomized clinical trials (RCTs). Due to the rise in the use of EHRs spurred by the HITECH Act and Affordable Care Act (ACA) and the availability of electronic health care data, there has been an increasing desire to use this data for regulatory purposes and there has been a rapid increase in the trend to include RWD and other non-interventional study designs in marketing applications. Most NDAs and BLAs that were approved in the past two years included at least one real world or non-interventional study. While study designs other than RCTs can and have been used to provide regulatory authorities with additional evidence of efficacy or safety, they currently present numerous challenges for both regulatory reviewers and sponsors. In our previous four papers (1,2,3,4), we examined the current regulatory landscape for submitting RWD, the gaps and limitations using the current data standards for submitting RWD, and the challenges of providing the traceability required for regulatory review. We offered recommendations for submitting RWD using the current standards. In this paper, we build on our previous work and demonstrate how our recommendations can be applied when submitting RWD as part of a marketing application. These recommendations are based on FDA's Real World Evidence (RWE) guidance, our previous papers, and how other common data models (OMOP, HL7 FHIR, and PCORnet) model RWD data elements. More specifically, we examine the updates that can be applied to a CDISC SDTM submission to better accommodate RWD. While our focus will be on exposure data, most of our recommendations can be applied to other domains or topic areas. For each submission challenge:

1) We examine the regulatory need and challenge that RWD presents.
2) We review how other data models such as HL7 FHIR, OMOP, and PCORnet handle a given domain or type of data.
3) We propose new or updated data elements and/or domains to more thoroughly represent RWD for a regulatory submission.

BACKGROUND

REAL WORLD EVIDENCE (RWE)

RWE can come from a variety of sources including electronic healthcare records (EHRs), payer administration claims data, 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 (5) 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 by Purpura in 2022 (6) estimated that 95% of approved NDAs and BLAs in 2021 included an RWE study.

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 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; these constructs may raise different types of considerations and these considerations may vary depending on the use of the RWE. For the purposes of this framework, FDA defines RWD and RWE as follows:

- **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 and RWE as the data derived from the RWD to be submitted to the FDA.

REGULATORY LANDSCAPE

The 21st Century Cures Act (7) enacted in 2016, and mandated that the FDA create a framework to evaluate RWE to support marketing applications for new drugs or new indications for drugs already approved. Subsequently, in 2018, the FDA published its framework for RWE, the *Framework for FDA'S Real World Evidence Program* (8). 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 will focus on these three areas when evaluating RWD submitted in support of a marketing application, just as they do for RCTs. The guidance notes there is no change in the standards for review of RWD versus RCT data. In 2021, the FDA expanded its program by issuing four more guidance documents that provided additional considerations, 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* (9) 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 (10).

In 2021, the FDA issued a second guidance, *Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry* (11), 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 (12). The primary takeaway from this guidance is that RWD and data from other non-interventional study designs must be submitted using the standards documented in the FDA Data Standards Catalog (13). To date, this means that RWD *must* be submitted using CDISC standards; thus, sponsors need to convert real-world datasets into CDISC SDTM and ADaM format when submitting this information in support of a marketing application. This guidance also provides advice on mapping RWD to CDISC format and considerations for the required documentation of this conversation. While the agency acknowledges that its current catalog of standards does not necessarily reflect data standards developed for real-world sources, it has indicated that it is considering updates, including recommendations for mapping. With respect to access to the source RWD, the guidance advises that sponsors confirm their ability to submit source patient-level data for any RWD used in studies submitted to the FDA in support of a marketing application.

In 2024, the FDA published the guidance on using RWD for marketing applications, *"Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry"* (14). This guidance covers a broad range of considerations for the use and submission of RWE. Especially pertinent, this guidance emphasizes the importance of characterizing the sources of RWD and the data elements used from these sources. The guidance makes it clear that the requirements for traceability for RWD are the same as those for RCT data. "Sponsors should address the procedures used to ensure completeness and accuracy of study data, as well as processes for data accrual, curation, and transformation over the data life cycle" including":

"Traceability of data elements to allow tracking of these elements back to their respective points of origin, with clear documentation of modifications that may have occurred." and

"Accuracy of mappings (e.g., in the presence of different coding systems, such as Systematized Nomenclature of Medicine—Clinical Terms [SNOMED CT] versus ICD-10-CM)."

The guidance also underscores the importance of maintaining the basic elements of good clinical practice, such a pre-specified analysis plan. The guidance proposes two sets of definitions be developed for key study variables, a conceptual and an operational definition and these be documented in the protocol. Key study variables are also specified and include study population inclusion and exclusion criteria, exposure details, outcomes, and covariates. The guidance also specifies that RWD be characterized for completeness, conformance, and plausibility of data values. Sponsors need to demonstrate that RWD is not altered meaningfully from the source data to the final data set analyzed and ultimately submitted to the FDA. The Data Management Process and Quality Management, including documentation of the Quality Assurance (QA) and Quality Control (QC) Plan should be submitted with the aim of facilitating FDA understanding of the data cleaning/curation process.

This guidance stresses the need for regulatory reviewers to understand each source of RWD and how the key data elements in the submission database can be traced back to the RWD source data. This guidance marks a step forward in integrating EHR and claims data into the regulatory framework of drug and biologic approvals.

CHALLENGES SUBMITTING RWD USING CDISC STANDARDS

This presentation follows four previous RWD papers by the same authors presented at PhUSE conferences since 2022. Our first paper focused on data from electronic health records (EHRs). We highlighted the challenges of using data not collected under a protocol and the issue of achieving data accuracy comparable to that of a RCT. We identified gaps in the current tabulation standard, CDISC, when used to represent RWE (see Table 1 below).

Table 1. Gaps in Current Submission Standards

Topic	RCT	RWD
Traceability	Annotated CRF /define file	Some type of traceability to source / annotated document describing source? Need to go back to original source
Visit Schedule	Protocol / SV domain	Need to determine appropriate encounters, type, length, location/site, type of site.
Exposure	Protocol determined	Reside in various record types. Need to determine what variables are needed to derive exposure, determine quality and certainty or records.
Trial Summary Dataset	TS	For RWE studies is TS needed? If so, are new parameters related to RWE needed.
Core Variables	CDISC has defined core variables for RCTs	Set of core variables for RWE has not been identified
Terminology	MedDRA / WHODrug	SNOMED / ICD. Need to represent multiple coding systems and their mappings.

Our next two presentations were a deeper dive into some of the challenges identified when submitting RWD to support a new marketing application and presented possible solutions. These challenge areas included cohort selection and bias identification, source data characterization, the lack of terminology harmonization between clinical research and healthcare, traceability, the temporality of data collection and differentiating between data collected directly from the EHR versus data added by the current investigator.

Our most recent presentation looked at future clinical data submission standards. We reviewed the landscape of existing standards for RWD/E from academia, government, industry, SDOs, the health technology industry and consortia. We identified 3 candidates for use with RWE: CDISC, FHIR, OMOP and considered options for implementation (single or hybrid model) and discussed the ability of each standard to represent the identified challenges for a RWD submission.

The landscape for submitting RWD is still evolving. FDA guidance documents cited above underscore comments from the Agency leadership that the data standards and documentation required for submitting RWD are the same as for submitting RCT data, which presents challenges for sponsors. As noted above, our previous four papers (1,2,3,4) examined the gaps and limitations for submitting RWD using the current data standards and the challenges of providing the traceability for regulatory review. These challenges are summarized in Table 2. Note that most of these challenges stem from the fact that RWD, by definition, is not collected under a protocol and there is no monitoring of the collection or documentation of data and by research study staff. In addition, the business cases for using RWD and data from RCTs differ. RWD, by definition, 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 use different data standards that contain different concept granularities and coding systems compared to those preferred by FDA, included in CDISC, and submitted for regulatory review.

Table 2. Traceability Challenges: RCTs vs RWD

RWD PRESENTS NEW TRACEABILITY CHALLENGES: RCTS VS RWD

Topic	RCT	RWD
Data Collection	Collected under a protocol	Collected in real world settings
Data Monitoring	Data monitored and cleaned	No monitoring or cleaning
Data Entry	Data collected via CRF	Data entered in EHR
Data Source	Data designed and collected by sponsor	Curated data acquired by sponsor from vendor
Availability or Source Data	Owned by sponsor	Owned by HCPs / Aggregators
Documentation	Protocol / CRFs / DMP	Aggregators
Data Uniformity	Uniform data entry across sites	Standards/formats differ by site
Terminologies/Vocabularies	Uniform across sites	May differ by site/type of RWD/region

Note: RCT (Randomized Clinical Trial)

ENVIRONMENTAL SCAN

In this section, we review common standards that are used to represent clinical data from both RCTs, RWD, and other study designs.

CDISC

Clinical Data Interchange Standards Consortium (CDISC) is a development effort that began in 1999 to standardize the data commonly collected for RCTs. CDISC has worked closely with both pharma and FDA to develop its standards to meet the needs of both stakeholders. Since December 2016, it has been recognized by FDA as the required standard for submitting clinical data to support a marketing application and is specified, along with other standards, in FDA's Data Standards Catalog. CDISC publishes a framework for both a tabulation model to represent collected data (i.e., the Study Data Tabulation Model (SDTM)) and an analysis model to represent analysis data (i.e., the Analysis Data Model (ADaM)) as well as associated metadata for these standards (Define.xml). The SDTM is organized into topic-related domains whereas the ADaM is organized into standard data structures for analysis purposes. CDISC datasets utilize CDISC Controlled Terminology where applicable, which is developed by CDISC in collaboration with the National Cancer Institute's Enterprise Vocabulary Services (EVS). It represents the set of codelists, and valid values used within CDISC-defined datasets, many of which are required to submit CDISC-compliant datasets to regulatory agencies.[1] MedDRA and WHODrug are both additional dictionaries leveraged in CDISC standards, and LOINC and SNOMED dictionaries may be used when applicable to the data and to the submission. CDISC contains other standards related to the clinical research process, including for non-clinical data (SEND), clinical study planning (Protocol), and clinical data

[1] More information is available at CDISC's Controlled Terminology webpage: https://www.cdisc.org/standards/terminology/controlled-terminology.

collection (CDASH). Further, data exchange standards are supported by CDISC such as dataset metadata submitted using the Define-XML and the ODM and the recently released Dataset-JSON standard.

Read more about CDISC

FHIR

Fast Healthcare Interoperability Resources (FHIR) is a standard for health care data exchange published by HL7 in 2012 and had begun adoption by EHRs starting with its mandate by the ASTP (ONC) in 2020. FHIR provides a framework for the exchange, integration, sharing, and retrieval of healthcare information across healthcare systems. The primary use cases for FHIR are to exchange healthcare and medical claims data. In 2019, over 80% of hospitals implemented API technology enabled with FHIR (15). Data elements in FHIR are grouped into resources in a manner that supports business objectives for medical health records and claims data. For those familiar with CDISC, resources are like domains. One key difference between the two standards is that FHIR allows resources to be linked directly to other resources without the use of additional "support" data (e.g., RELREC dataset).
Read more about HL7 FHIR

OMOP

The Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM) is an open community data standard, designed to standardize the structure and content of observational healthcare data. It was originally developed under a grant from NIH to enable the merging of EHR and claims data and facilitate efficient analyses that can produce reliable evidence. It was designed to address the challenges associated with the heterogeneity of healthcare data by providing a standardized structure for organizing and representing diverse observational databases. It is currently maintained by Observational Health Data Sciences and Informatics (OHDSI) consortium.

At its core, OMOP CDM is designed to harmonize disparate healthcare data sources, ranging from EHRs and claims data to registries and other real-world evidence. The primary objective is to create a consistent framework that facilitates the integration and analysis of data from different healthcare systems and settings. By establishing a common language, researchers and analysts can navigate through a wealth of information with greater efficiency, promoting cross-study collaboration and data sharing.

Designed to focus on observational data and reflecting real-world clinical practices and patient outcomes, it stands in contrast to the CDISC data standard which was developed for representing data from clinical trials with controlled environments and selected patient populations. By leveraging observational data, the OMOP CDM is designed to enable researchers to gain insights into the effectiveness and safety of medical treatments, patient characteristics, and disease patterns in diverse and representative populations.

The model comprises a set of standardized tables and concepts, providing a consistent structure for storing healthcare data elements such as patients, diagnoses, medications, and procedures. This standardization facilitates the translation of complex clinical information into a format that is easily accessible and analyzable. Moreover, OMOP CDM incorporates standardized terminologies, such as SNOMED CT and LOINC, and seeks to ensure semantic consistency to enhance interoperability across the various systems housing healthcare data.

A central component of the OMOP CDM is the OHDSI standardized vocabularies. The OHDSI vocabularies facilitate organization and standardization of medical terms used across the clinical domains of the OMOP CDM and enable standardized analytics across the database when constructing exposure and outcome phenotypes and other features within characterization, population-level effect estimation, and patient-level prediction studies.

OMOP CDM supports large-scale collaborative research efforts and enables the development of advanced analytical methods and tools within the OHDSI community. OHDSI offers a wide range of open-source tools designed for use with one or more of the databases of the CDM to support an array of data-analytics use cases on observational patient-level data. These tools facilitate performing analyses by filling in standard templates improving the reproducibility and transparency of analyses by using standard algorithms to compute commonly used metrics, such as incidence rate. OMOP also provides tools to evaluate the quality of data formatted in the OMOP CDM. They use a framework established by Kahn et al. which defines data quality (DQ) as consisting of conformance, completeness, and plausibility. More on OMOP's data quality can be found in the The Book Of OHDSI, Chapter 15 (16).

Read more about the OMOP Common Data Model
Read more about OHDSI's standardized vocabularies

PCORNET Common Data Model (CDM)

The Patient-Centered Outcomes Research Institute (PCORI), authorized by Congress in 2010, is an independent, nonprofit research organization that funds patient-centered comparative clinical effectiveness research (CER). The PCORnet CDM is a data standards specification that defines the organization and representation of data for the PCORnet Distributed Research Network (DRN). The PCORnet DRN allows multi-site patient–centered research across the Clinical Research Networks (CRNs) and other interested contributors. Like Sentinel and OHDSI, the PCORnet distributed network facilitates the conduct of observational research while allowing each participating organization to maintain physical and operational control over its own data. The PCORnet DRN operations center and infrastructure, including the CDM, is led by the PCORnet Coordinating Center and overseen by the governance established by PCORnet's stakeholders.

The PCORnet CDM, first released in 2014, was originally based on the Mini-Sentinel Common Data Model v4.0 (MSCDM v4.0). It has evolved over time informed, in part, by the development of other distributed initiatives such as the HMO Research Network, the Vaccine Safety Datalink, various AHRQ Distributed Research Network projects, and the ONC Standards & Interoperability Framework Query Health Initiative. The PCORnet CDM uses standard healthcare terminologies including ICD, SNOMED, CPT, HCPCS, and LOINC® to enable interoperability. The most recent version of the CDM, V6.1, was released in 2023. Version 4.0 of the PCORnet CDM introduced the concept of expanded value sets for fields with dozens or hundreds of allowable options, some related to exposure including DOSE_FORM, ROUTE, and UNIT as well as others such as LAB_RESULT_CM and RESULT_UNIT.

All data available through PCORnet are subject to a two-stage quality assessment process. The first stage checks data for conformance to ensure it adheres to the CDM and checks for completeness, including that diagnosis codes are aligned, plausible, and persistent, to ensure that records are maintained between refreshes. A list of all checks is provided on the website: https://pcornet.org/wp-content/uploads/2024/08/PCORnet-Data-Checks-v16.pdf. Thoughtfully chosen data quality metrics provided with RWE (and possibly for the original source of RWD as well) could help FDA understand the data they are reviewing more quickly. For example, it could show if missing values appear to be missing at random or if there is a pattern to missingness. It could also show if the data selected differs in quality from the source data as a whole. Data quality itself is a large topic that is not covered in detail in this paper but may possibly be covered in a subsequent presentation.

Read more about the PCORnet.
Read more about PCORnet data.

EXPOSURE DATA CONSIDERATIONS

In this section we focus on representing exposure data for submission. We define exposure as the protocol-specified study treatment. We examine the challenges of submitting exposure data, how each of the data models presented above represent exposure data, and provide recommendations based on these data models to address these challenges.

Exposure data is key to providing evidence of the benefits or risks of a treatment. As stated in the table above, the study treatments as well as dosage and frequency, are well characterized in the protocol for RCT data. Exposure data is collected via a CRF that is standardized across sites. Furthermore, exposure data is thoroughly monitored and cleaned. In contrast, exposure data from RWD can come from a variety of sources which are not uniform across sites. These data are not monitored or cleaned, and in many cases are incomplete. In addition, the business cases for using RWD and RCTs differ. RWD is typically 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 that contain different concepts and coding systems compared to those defined within CDISC and used in regulatory review.

CDISC standards were designed and optimized to represent RCT data. Not surprisingly, CDISC SDTM domains (EX and EC) designed to submit exposure data are not structured to adequately represent exposure data coming from RWD sources. CDISC exposure domains also lack many of the data elements/concepts needed to capture RWD. On the other hand, common data models such as HL7 FHIR, OMOP, and PCORnet were designed to represent EHR/claims and observational study designs and also do not have all the data elements needed for FDA submission.

The CDISC SDTM contains two domains for submitting exposure data: Exposure (EX) and Exposure as collected (EC). Figure 1 abstracted from the CDISC SDTMIGv3.4 illustrates the CDISC approach to modeling exposure data. As shown in Figure 1, the focus of CDISC exposure domains is on clinical trials. While CDISC was optimized to represent RCT data for medical products whose safety and efficacy has not yet been established. In contrast, data models such as HL7 FHIR, OMOP, and PCORnet were designed for a different business case: to represent EHR/claims and observational studies data. Exposure in the case of routine healthcare and observation studies means exposure to

medical products that have already been approved and have an established efficacy and a known safety profile for the population that it was tested in.

Figure 1. CDISC SDTMIG3.4 Exposure Domains

6.1.3 Exposure Domains

Clinical trial study designs can range from open label (where subjects and investigators know which product each subject is receiving) to blinded (where the subject, investigator, or anyone assessing the outcome is unaware of the treatment assignment(s) to reduce potential for bias). To support standardization of various collection methods and details, as well as process differences between open-label and blinded studies, 2 SDTM domains based on the Interventions General Observation Class are available to represent details of subject exposure to protocol-specified study treatment(s): Exposure (EX) and Exposure as Collected (EC).

6.1.3.1 Exposure (EX)

EX – Description/Overview

An interventions domain that contains the details of a subject's exposure to protocol-specified study treatment. Study treatment may be any intervention that is prospectively defined as a test material within a study, and is typically but not always supplied to the subject.

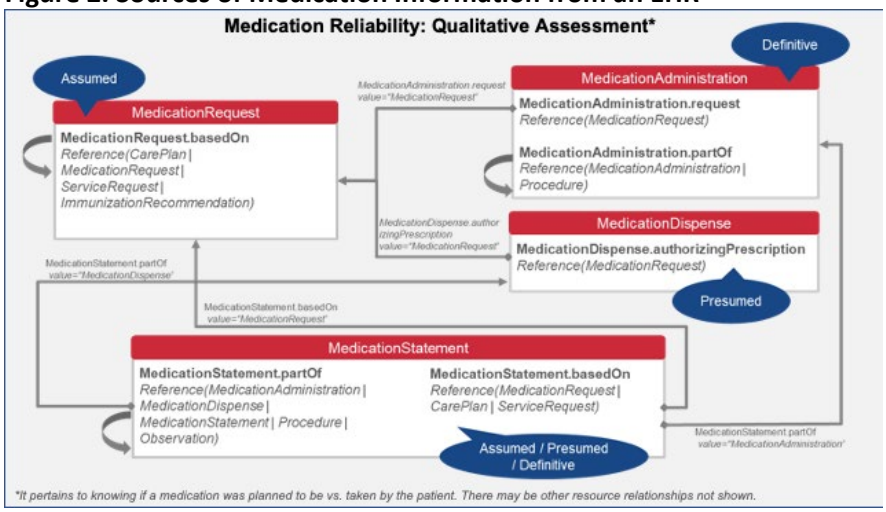
6.1.3.2 Exposure as Collected (EC)

EC – Description/Overview

An interventions domain that contains information about protocol-specified study treatment administrations, as collected.

Exposure/treatment data in EHRs typically reside in four distinct record types: 1) medication request or medication ordered records; 2) medication dispensed records; 3) medication administered records (medication administered at a clinical site); and 4) medication statement records (medication usage reported by a patient or significant other). EHRs may or may not contain medication data from one or more of these sources. Figure 2 from the HL7 FHIR implementation guide "Retrieval of Real-World Data for Clinical Research" (17) illustrates the various sources of medication information as well as the reliability of each source.

Figure 2. Sources of Medication Information from an EHR



While data must be submitted using CDISC standards, several concepts from both HL7 FHIR, OMOP, and PCORnet can be utilized when submitting exposure data to more accurately represent exposure data and provide the traceability required for regulatory review. In addition to the four resources (or 'domains' in CDISC) that FHIR has to model data for each of the record types noted above, FHIR also contains a "Medication" resource to characterize a medication.

The OMOP data model represents exposure data in the DRUG_Exposure Table. This table contains a data element DRUG_TYPE_ CONCEPT_ID, which provides information about the provenance of the record, as well as data elements to represent both the source, standardized code, and coding system for a given medication.

Like HL7 FHIR, the PCORnet standard recognizes that there are various types of exposure records. PCORnet contains three tables for modeling exposure data: 1) Dispensing – prescriptions filled through a community, mail-order or

hospital pharmacy; 2) Prescribing – provider orders for medication dispensing and/or administration; and 3) Med-Admin – records of medications administered to patients by healthcare providers. Furthermore, PCORnet contains data elements both to store concepts as captured from the source and data elements standardized to meet submission requirements.

RECOMMENDATIONS

Since CDISC is currently required by FDA for data supporting new marketing applications, we recommend submitting the collected data from all available sources in a format as "SDTM" as possible. This can mean either creating a custom exposure domain(s) or adding as many supplemental variables as needed to the SDTM EC and/or EX domains. Then, we can draw upon the FHIR HL7, OMOP, and PCORnet data models to improve the representation, traceability, and provenance needed for regulatory review of RWD.

DATA PROVENANCE: RECORD TYPES

The FDA guidance cited above, *"Real-World Data: Assessing Electronic Health Records and Medical Claims Data To support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry"* (14), recommends submitting documentation including "traceability of data elements to allow tracking of these elements back to their respective points of origin, with clear documentation of modifications that may have occurred". However, within the CDISC SDTM EX and EC domains, no details exist to indicate the provenance of exposure data since data collection in RCTs is traced using a CRF. That is, there is no ability to record whether an exposure record is coming from an administration record, a medication statement, a dispense record, or a prescription record. Drawing on FHIR, OMOP, and PCORnet we suggest two possible approaches: 1) as done in OMOP, expanding the EX/EC domains to include a data element for record type; or 2) model exposure data in multiple domains like FHIR or PCORnet, and create a custom domain for each record type. In ADaM, exposure to a given treatment can be constructed from all available sources as specified in the statistical analysis plan. Furthermore, a derived "confidence" variable can be useful to report the level of confidence in the accuracy of each source containing exposure data.

TRACEABILITY

As noted above, FDA guidance stresses the importance of traceability for regulatory review. The SDTM domains EX/EC only contain a single data element to document the study treatment (EXTRT or ECTRT, respectively). This may suffice for RCTs, as the study treatment and its administration are well documented in the protocol and is carefully controlled during the clinical trial. However, for RWD, regulatory reviewers are interested in how the study treatment was represented in the source data because that gives an indication of the certainty of the exposure. Figure 3 demonstrates how a medication from an EHR record is represented in FHIR. As shown in Figure 3, three data elements are used: 1) Display – a text description of the medication; 2) Code – a code that identifies the medication; and 3) Coding system – the coding system associated with the code.

Figure 3. Representing a Medication in FHIR

```
{
  "resourceType" : "MedicationRequest",
  "id" : "medrx0333",
  "contained" : [{
    "resourceType" : "Medication",
    "id" : "med0310",
    "code" : {
      "coding" : [{
        "system" : "http://snomed.info/sct",
        "code" : "430127000",
        "display" : "Oxycodone-containing product in oral dose form"
      }]
    }
  }],
```

Figure 4 presents the PCORnet solution for capturing a medication name. Each exposure related table has one or more data elements for capturing the source medication and and/or related code (as well as the standardized name and

code). All three tables in Figure 4 have a field to capture the medication code from the source. Two of the tables also contain data elements to capture full text description of the medication from the source. The PCORnet data model employs a general convention for "raw data fields": 'RAW' will be part of the data element name (i.e., RAW_NDC). Using this naming convention, PCORnet also contains data elements (not shown here) to represent both the source and standardized values for several other exposure related concepts such as route, dose amount, and dose units.

Figure 4. Representing a Medication in PCORnet

Table	Concept	Note
DISPENSING	RAW_NDC	Field for originating value, prior to mapping into the PCORnet CDM value set.
PRESCRIBING	RAW_RX_MED_NAME	Field for originating, full textual medication name from the source.
PRESCRIBING	RAW_RX_NDC	Field for originating value, prior to mapping into the PCORnet CDM value set
MED_ADMIN	RAW_MEDADMIN_MED_NAME	Field for originating, full textual medication name from the source.
MED_ADMIN	RAW_MEDADMIN_CODE	Field for originating value, prior to mapping into the PCORnet CDM value set.

Like PCORnet, OMOP also contains a data element (DRUG_SOURCE_VALUE) to house the verbatim value from the source data. This is typically the source code for the medication as it appears in the source data. OMOP also has a related data element called "DRUG_SOURCE_CONCEPT_ID". This field stores an ID which maps to a coding system (such as NDC) and a text description.

Based on the implementation in PCORnet with input from OMOP and FHIR, we recommend expanding EX/EC to include additional variables to describe study treatment in the source data: 1) Source Medication Name; 2) Source Medication Code; and 3) Source Medication Coding System. An example of how this can be implemented is shown in Figure 5. This solution can be expanded to include other data elements characterizing study treatment such as, but not limited to, route, dose, and dose units. Furthermore, we recommend implementing this strategy (collecting the verbatim name or term from the source data along with the associated code and coding system) in all applicable domains to enhance traceability. This solution is especially relevant for specimen-based Lab Findings domains, Vital Signs, and Adverse Events.

Figure 5. Expanding EX/EC to include additional source variables.

Source Medication Name	Source Medication Code	Source Medication Coding System	Standardized Name of Treatment
Oxycodone-containing product in oral dose form	430127000	http://snomed.info/sct"	Oxycodone

Does not currently exist in SDTMIG

EXTRT

ADDITIONAL DOSING INFORMATION

While exposure is well characterized in RCTs, exposure in RWD is constructed from the various sources described above. In many cases, only one or two of these sources will be available. Furthermore, these data may contain a significant amount of missing data. Therefore, we recommend adding several data elements to the CDISC SDTM EX/EC domains that are often contained in RWD to facilitate constructing an individual subject's exposure to the study treatment. Again, we can consult other CDMs that are optimized for RWD to achieve this goal. Figures 6, 7, and 8 provide extracts from the FHIR Medication Request and Medication Dispense resources, the PCORnet Dispensing and Prescribing tables, and the OMOP Drug_Exposure table. While the CDISC EX/EC domains do not contain concepts for "days supply", "fill quantity", or the "total number of refills", FHIR, PCORnet, and OMOP all contain data elements representing these concepts. We strongly recommend adding these data variables to EX/EC. Since end date may be missing from exposure data within RWD, these three data elements can be especially useful in deriving end date when

it is missing. In addition, Figure 9 provides an example of these data elements within a prescription record for a medication taken by the subject deemed to be a 'concomitant medication' within the example RWD study.

Figure 6. HL7 FHIR, Additional Dosing Variables

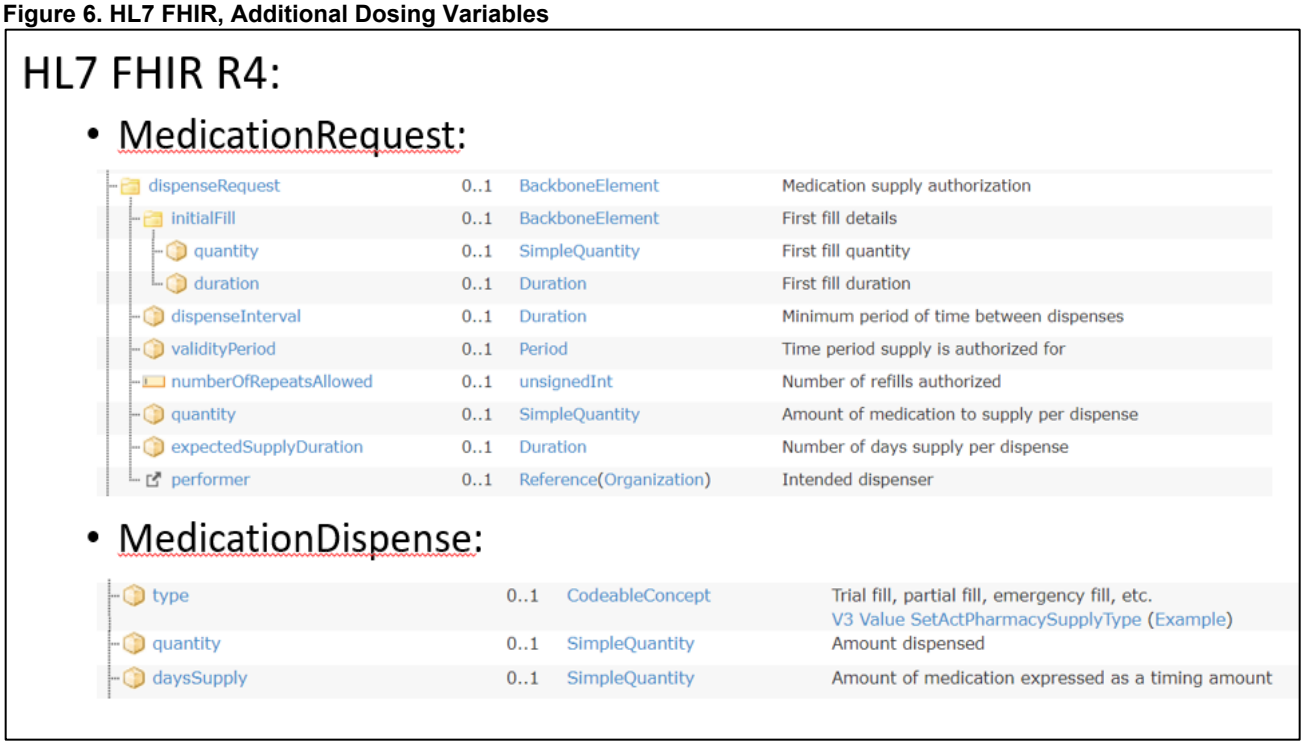


Figure 7. PCORnet, Additional Dosing Variables

PCORnet v6.1:

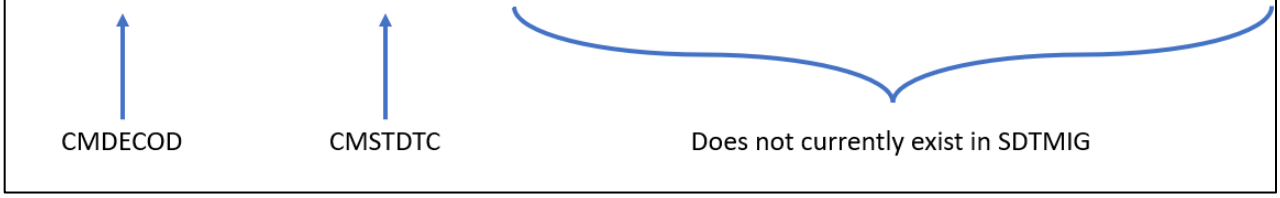
Table	Concept	Note
DISPENSING	DISPENSE_SUP	Days supply. Number of days that the medication supports based on the number of doses as reported by the pharmacist. This amount is typically found on the dispensing record. Integer values are expected.
DISPENSING	DISPENSE_AMT	Number of units (pills, tablets, vials) dispensed. Net amount per NDC per dispensing. This amount is typically found on the dispensing record. Positive values are expected.
PRESCRIBING	RX_QUANTITY	Quantity ordered.
PRESCRIBING	RX_REFILLS	Number of refills ordered (not including the original prescription). If no refills are ordered, the value should be zero.
PRESCRIBING	RX_DAYS_SUPPLY	Number of days supply ordered, as specified by the prescription.

Figure 8. OMOP, Additional Dosing Variables

REFILLS	NULL	The number of automatic refills after the initial prescription that are part of the prescription system in many countries. The initial prescription is not counted, values start with NULL. In the case of Lauren's acetaminophen she did not have any refills so the value is NULL.
QUANTITY	60	The quantity of drug as recorded in the original prescription or dispensing record.
DAYS_SUPPLY	30	The number of days of supply of the medication as prescribed.

Figure 9. Example of Adding Additional Data Elements to a Prescription Record

Standardized Medication Name	Start Date/Time of Medication	Days Supply	Fill Quantity	Total Number of Refills
Amoxicillin	2024-09-10	14	28	0



CONTEXTUAL VARIABLES

In RCTs, the protocol specifies *how* an assessment is conducted, *who* performs the assessment, and *where* the assessment is completed. As a result, this information is usually not captured during data collection, nor has it been required by reviewers. Since RWD is not collected under a protocol, submitting this additional information surrounding the context of an assessment can help increase a reviewer's confidence and understanding of the submitted RWD.

Where applicable, we recommend adding supplemental "contextual" variables as additional qualifiers to the SDTM EX/EC domains. These can include who conducted the assessment and the name and location of the site where the assessment was done. The rationale for providing these contextual variables is that all sites are not equal (e.g., a large medical center versus a corner drug store care clinic) and that clinical staff have different skill sets, experience, and specializations. FHIR, OMOP, and PCORnet take a similar approach to providing this information; medication related resources or tables contain data elements for Provider ID, Encounter ID, and Facility or Location ID. These data elements link to either a Provider table, Encounter table, or a Facility/Location table (or resource) that contains extensive information about a given provider, encounter, or facility. For example, the Prescribing Table in PCORnet contains the variables RX_PROVIDERID and ENCOUNTERID. The RX_PROVIDERID field stores a code for the provider who prescribed the medication and links to the Provider table, which contains extensive details about the provider. Similarly, the ENCOUNTERID field holds a code for the encounter associated with the prescribing event and links to the Encounter table, which provides details about both the facility and encounter. OMOP and FHIR use the same methodology for linking a medication event with a provider, facility, and encounter. An alternative to this could be moving the FDA Bioresearch Monitoring (BIMO) data into the CDISC model, but it would need to account for multiple facilities and providers associated with a single patient, where it is currently designed for RCTs where a patient visits a single site for all care associated with the RCT.

To facilitate regulatory review, we recommend implementing the FHIR/PCORnet/OMOP strategy:

1) Add Provider ID and Facility ID to the SDTM EX/EC domains (as well as other applicable domains) as supplemental qualifiers to the parent domain(s).
2) Create a PROVIDER domain as a custom domain to store details about a given provider.

3) Either expand the Subject Visits (SV) domain by adding facility related data elements or create a new domain which contains details about a given facility.
4) Provider ID will link EX/EC to the PROVIDER domain.
5) Facility ID will link EX/EC to the domain storing the facility information.

OTHER VARIABLES

Additional site, subject, and record identifiers can facilitate traceability to the source data and provide increased confidence in RWD to regulatory reviewers. These data elements are especially needed for RWD, since records can come from a variety of sources. Therefore, we recommend adding three supplemental variables to the exposure domain(s) as well as to all other domains containing subject data: 1) RWD source – This indicates where the RWD comes from and can have values like EHR, Registry, Claims, etc.; 2) Source subject identifier – The subject identifier contained in the source data. This will likely be encrypted or scrambled; and 3) Source record identifier – This is the record identifier from the source data (also likely encrypted) that can be used to link back to the original record in the source data. In many cases, this will also be the encounter identifier for a subject.

CONCLUSIONS

There is a vast and growing amount of healthcare data exchanged in a variety of formats. This data already assists the FDA in post-market assessment of approved drugs but the increasing volume and variety of data has the potential to provide information about the potential benefits and risks of a medical product used off-label and to serve as a control group for new medical product approvals. However, this data available from non-RCTs results in data collected using both different data standards and different file formats and the lack of monitoring of the data collection and documentation can lead to quality issues which may impact the evaluation of safety and efficacy, particularly if the study population is small. We chose to highlight the issues associated with exposure in this paper as exposure of a patient to a medication is necessary to ensure before other elements related to drug safety and efficacy can be assessed accurately. Although not covered in this paper, it is worth reminding readers that without the use of randomization to select patients, which so far has not been seen in the use of RWE submitted to the FDA. Thus, the risk of bias in the RWE compounds the issues related to assessing RWE.

Due to the mandates of the 21st century cures act, as well as the desire to learn from patients' experiences that are documented in the various sources of RWD, there is an expectation from industry that RWD should be included in marketing applications for medical products. It is incumbent on sponsors, regulatory authorities, standards development organizations, and other industry-related groups to collaborate and agree upon the data elements needed to ensure an accurate efficacy and safety profile of a drug based on RWD, and then to design data standards that will allow for more efficient and effective regulatory review of real-world data. Whether the future is a single standard or a hybrid approach, regulatory reviewers are tasked with ensuring the safety of the public and require submission standards that allow them to understand the certainty of exposure as well as other information which contributes to making an accurate regulatory decision. There are lessons to be taken from the data models used for epidemiology and post market studies. However, a pre-market review of a new drug demands more rigor, and the data standards need to allow for this.

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) Ferko, S., Holt, I., & Abolafia, J., (2023). “Future Clinical Data Submission Standards: CDISC, FHIR, OMOP, or Hybrid Model”. Paper presented at the PHUSE US Annual Conference 2024, Bethesda, MD.
(5) Aetion. “Aetion_eBook_2021_The role of RWE in FDA Approvals”. https://aetion.com/evidence-hub/welcome-to-the-evidence-hub/
(6) 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.
(7) 21st Century Cures Act. (2016). https://www.congress.gov/114/plaws/publ255/PLAW-114publ255.pdf
(8) FDA Guidance (2018). “Framework for FDA’s Real-World Evidence Program”. December 2018. https://www.fda.gov/media/120060/download
(9) 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
(10) FDA Guidance. (2023). “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. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/considerations-use-real-world-data-and-real-world-evidence-support-regulatory-decision-making-drug
(11) 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
(12) FDA Guidance. (2023). “Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry". December 2023. https://www.fda.gov/media/153341/download
(13) FDA Guidance. (2023). “Data Standards Catalog”. April 2023. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/data-standards-catalog
(14) FDA Guidance. (2024). "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. https://www.fda.gov/media/152503/download
(15) Posnack, Steven and Barker, W. (2021). “The Heat is On: US Caught FHIR in 2019”. The Heat is On: US Caught FHIR in 2019 - Health IT Buzz
(16) The Book of OHDSI (2021). https://ohdsi.github.io/TheBookOfOhdsi/
(17) HL7 Implementation Guide (2023). “Retrieval of Real World Data for Clinical Research”. May 2023. https://build.fhir.org/ig/HL7/vulcan-rwd/medications.html

CONTACT INFORMATION

Jeffrey Abolafia
Certara/Pinnacle 21
mailto:jeffrey.abolafia@certara.com