
Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry

DRAFT GUIDANCE

This guidance document is being distributed for comment purposes only.

Comments and suggestions regarding this draft document should be submitted within 60 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance. Submit electronic comments to <https://www.regulations.gov>. Submit written comments to the Dockets Management Staff (HFA-305), Food and Drug Administration, 5630 Fishers Lane, Rm. 1061, Rockville, MD 20852. All comments should be identified with the docket number listed in the notice of availability that publishes in the *Federal Register*.

For questions regarding this draft document or the Real-World Evidence Program, please email CDERMedicalPolicy-RealWorldEvidence@fda.hhs.gov.

**U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)**

**October 2021
Real-World Data/Real-World Evidence (RWD/RWE)**

Data Standards for Drug and Biological Product Submissions Containing Real-World Data Guidance for Industry

Additional copies are available from:
Office of Communications, Division of Drug Information
Center for Drug Evaluation and Research
Food and Drug Administration
10001 New Hampshire Ave., Hillandale Bldg., 4th Floor
Silver Spring, MD 20993-0002
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353
Email: druginfo@fda.hhs.gov
<https://www.fda.gov/drugs/guidance-compliance-regulatory-information/guidances-drugs>

and/or

Office of Communication, Outreach and Development
Center for Biologics Evaluation and Research
Food and Drug Administration
10903 New Hampshire Ave., Bldg. 71, Room 3128
Silver Spring, MD 20993-0002
Phone: 800-835-4709 or 240-402-8010
Email: ocod@fda.hhs.gov
<https://www.fda.gov/vaccines-blood-biologics/guidance-compliance-regulatory-information-biologics/biologics-guidances>

U.S. Department of Health and Human Services
Food and Drug Administration
Center for Drug Evaluation and Research (CDER)
Center for Biologics Evaluation and Research (CBER)

October 2021
Real-World Data/Real-World Evidence (RWD/RWE)

Contains Nonbinding Recommendations

Draft — Not for Implementation

TABLE OF CONTENTS

| | | |
|-------------|--|----------|
| I. | INTRODUCTION AND SCOPE | 1 |
| II. | REGULATORY BACKGROUND..... | 2 |
| III. | APPLYING CURRENTLY SUPPORTED DATA STANDARDS TO STUDY DATA DERIVED FROM REAL-WORLD DATA SOURCES..... | 3 |
| A. | Challenges in Real-World Data Standardization..... | 3 |
| B. | Documentation of Processes for Managing Real-World Data | 3 |
| C. | Considerations for Conforming Real-World Data to Currently Supported FDA Study Data Standards..... | 3 |
| D. | Considerations for Mapping Real-World Data to Study Data Submission Standards | 4 |
| E. | Considerations for Data Transformations..... | 5 |
| IV. | GLOSSARY..... | 5 |
| | APPENDIX: EXAMPLES OF MAPPING HEALTH CARE DATA TO CDISC SDTM | 8 |

Contains Nonbinding Recommendations

Draft — Not for Implementation

1 **Data Standards for Drug and Biological Product Submissions**
2 **Containing Real-World Data**
3 **Guidance for Industry¹**
4

5
6 This draft guidance, when finalized, will represent the current thinking of the Food and Drug
7 Administration (FDA or Agency) on this topic. It does not establish any rights for any person and is not
8 binding on FDA or the public. You can use an alternative approach if it satisfies the requirements of the
9 applicable statutes and regulations. To discuss an alternative approach, contact the FDA staff responsible
10 for this guidance as listed on the title page.
11

12
13
14
15 **I. INTRODUCTION AND SCOPE**

16
17 **21st Century Cures Act and Real-World Data**
18

19 The 21st Century Cures Act,² signed into law on December 13, 2016, is intended to accelerate
20 medical product development and bring innovations faster and more efficiently to the patients
21 who need them. Among other provisions, the 21st Century Cures Act added section 505F to the
22 Federal Food, Drug, and Cosmetic Act (FD&C Act) (21 U.S.C. 355g). Pursuant to this action,
23 calling for FDA to issue guidance on the use of real-world evidence (RWE) in regulatory
24 decision-making, FDA has created a framework for a program to evaluate the potential use of
25 real-world data (RWD) to generate RWE to help support the approval of new indication(s) for
26 drugs³ already approved under section 505(c) of the FD&C Act (21 U.S.C. 355(c)) or to help
27 support or satisfy post-approval study requirements (RWE Program).⁴
28

29 This guidance provides recommendations to sponsors for complying with section 745A(a) of the
30 FD&C Act (21 U.S.C. 379k-1(a)) when submitting RWD as study data in applicable drug
31 submissions. FDA is issuing this guidance as part of its RWE Program and to satisfy, in part, the
32 mandate under section 505F of the FD&C Act (21 U.S.C. 355g) to issue guidance on the use of
33 RWE in regulatory decision-making.⁵

¹ This guidance has been prepared by the Center for Drug Evaluation and Research (CDER) in cooperation with the Center for Biologics Evaluation and Research (CBER) at the Food and Drug Administration.

² Public Law 114-255.

³ For the purposes of this guidance, all references to *drugs* include both human drugs and biological products.

⁴ See *Framework for FDA's Real-World Evidence Program* (December 2018), available at <https://www.fda.gov/media/120060/download>. The framework and RWE Program also cover biological products licensed under the Public Health Service Act.

⁵ See section 505F(e) of the FD&C Act.

Contains Nonbinding Recommendations

Draft — Not for Implementation

34 This guidance addresses considerations for the use of ***data standards***⁶ currently supported by
35 FDA in applicable drug submissions containing study data⁷ derived from RWD sources. For the
36 purposes of this guidance, FDA defines *RWD* as data relating to individual patient health status
37 or the delivery of health care routinely collected from a variety of sources. Examples of RWD
38 include data from ***electronic health records*** (EHRs); ***medical claims data***, data from product and
39 disease ***registries***; patient-generated data (including data from in-home-use settings); and data
40 gathered from other sources that can inform on health status, such as mobile devices.

41
42 The contents of this document do not have the force and effect of law and are not meant to bind
43 the public in any way, unless specifically incorporated into a contract. This document is
44 intended only to provide clarity to the public regarding existing requirements under the law.
45 FDA guidance documents, including this guidance, should be viewed only as recommendations,
46 unless specific regulatory or statutory requirements are cited. The use of the word *should* in
47 FDA guidance means that something is suggested or recommended, but not required.

48
49

II. REGULATORY BACKGROUND

50
51
52 Under section 745A(a) of the FD&C Act, at least 24 months after the issuance of a final
53 guidance document in which FDA has specified the electronic format for submitting certain
54 submission types to the Agency, such content must be submitted electronically and in the format
55 specified by FDA.⁸ The guidance for industry, *Providing Regulatory Submissions In Electronic*
56 *Format — Standardized Study Data* (Study Data Guidance), and the technical specifications
57 referenced therein describe electronic submission requirements under section 745A(a) of the
58 FD&C Act for clinical and nonclinical study data contained in new drug applications (NDAs),
59 abbreviated new drug applications (ANDAs), certain biologics license applications (BLAs), and
60 certain investigational new drug applications (INDs) submitted to the Center for Drug Evaluation
61 and Research or the Center for Biologics Evaluation and Research.⁹ Given that these electronic
62 submission requirements apply to study data submitted in the covered application types, they
63 apply to RWD that is submitted as study data in such applications. That is, RWD submitted as
64 study data to NDAs, ANDAs, certain BLAs, and certain INDs, as further described in section
65 II.A of the Study Data Guidance, must be in an electronic format that the Agency can process,
66 review, and archive, unless such submission is exempt from the electronic submission
67 requirements or if FDA has granted a waiver.¹⁰ Currently, as stated in the Study Data Guidance,

⁶ See the Glossary (section VII) for definitions of words and phrases that are in bold italics at first mention throughout this guidance.

⁷ See FDA guidance for industry *Providing Regulatory Submissions in Electronic Format — Standardized Study Data* (June 2021). We update guidances periodically. For the most recent version of a guidance, check the FDA guidance web page at <https://www.fda.gov/regulatory-information/search-fda-guidance-documents>.

⁸ For additional information on how FDA interprets and intends to implement the electronic submission requirements of section 745A(a) of the FD&C Act, see guidance for industry *Providing Regulatory Submissions in Electronic Format – Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act* (December 2014).

⁹ See section II of the Study Data Guidance for more information on the types of submissions subject to electronic submission requirements for standardized study data and what submissions are exempt from such requirements.

¹⁰ Sponsors or applicants may apply for a waiver from the requirement to use specific versions of FDA-supported standards for the submission of study data using the waiver request process described in section II.D of the Study Data Guidance.

Contains Nonbinding Recommendations

Draft — Not for Implementation

68 the Agency can process, review, and archive electronic submissions of clinical and nonclinical
69 study data (including those derived from RWD sources) that use the standards specified in the
70 Data Standards Catalog (Catalog).¹¹ As that guidance explains, the Catalog provides a listing of
71 currently supported¹² and/or required standards, their uses, the date FDA will begin (or has
72 begun) to support a particular standard, the date such support ends (or will end), the date the
73 requirement to use a particular standard will begin (or has begun), the date such requirement
74 ends (or will end), and other pertinent information. FDA is issuing this guidance to provide
75 recommendations to sponsors for complying with section 745A(a) of the FD&C Act using
76 standards specified in the Catalog when submitting study data derived from RWD sources in
77 applicable drug submissions.

78
79

III. APPLYING CURRENTLY SUPPORTED DATA STANDARDS TO STUDY DATA DERIVED FROM REAL-WORLD DATA SOURCES

81
82
83

A. Challenges in Real-World Data Standardization

84

85 FDA recognizes the challenges involved in standardizing study data derived from RWD sources
86 for inclusion in applicable drug submissions. These challenges include but are not limited to: (1)
87 the variety of RWD sources and their inconsistent formats (e.g., EHR, registry); (2) the
88 differences in *source data* captured regionally and globally using different standards,
89 *terminologies*, and *exchange formats* for the representation of the same or similar data
90 elements¹³; (3) a wide range of methods and algorithms used to create datasets intended to
91 aggregate data; and (4) the many aspects of health care data that can affect the overall quality of
92 the data, including business processes and database structure, inconsistent vocabularies and
93 coding systems, and de-identification methodologies used to protect patient data when shared.

94

B. Documentation of Processes for Managing Real-World Data

95

96
97 During *data curation* and *data transformation*, adequate processes should be in place to increase
98 confidence in the resultant data. Documentation of these processes may include but are not
99 limited to electronic documentation (i.e., metadata-driven audit trails, quality control procedures,
100 etc.) of data additions, deletions, or alterations from the source data system to the final study
101 analytic data set(s). Sponsors should also document in their applicable drug submission changes
102 to data to conform to the current FDA-supported data standards, and the potential impacts of
103 these changes.

104

C. Considerations for Conforming Real-World Data to Currently Supported FDA Study Data Standards

105

106
107
108 FDA plans to issue further guidance and/or to update the Catalog with standards for study data
109 that are derived from RWD sources. Currently, and absent a waiver, sponsors submitting clinical

¹¹ The Catalog is available at <http://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm>.

¹² For the purposes of this document, “supported” means the receiving Center has established processes and technology to support receiving, processing, reviewing, and archiving files in the specified standard.

¹³ See data element at https://csrc.nist.gov/glossary/term/data_element.

Contains Nonbinding Recommendations

Draft — Not for Implementation

110 and nonclinical study data (including those derived from RWD sources) in submissions subject
111 to section 745A(a) of the FD&C Act are required to use the formats described in the Study Data
112 Guidance and the supported study data standards listed in the Catalog. Sponsors should refer to
113 the specifications, recommendations, and general considerations provided in the *Study Data*
114 *Technical Conformance Guide*¹⁴ when submitting study data in an applicable drug submission to
115 FDA. When seeking to conform RWD to data standards supported by FDA, sponsors should
116 consider the relevant data transformations, conversions, or *mappings* that may be needed to
117 produce study datasets in the required format in an applicable drug submission.

118
119 Sponsors should discuss early, with the appropriate FDA review division, any planned
120 submission of study data derived from RWD sources in an applicable drug submission and their
121 approaches for transforming the data to the current FDA-supported data standards. Sponsors
122 should describe these approaches, including in the protocol, data management plan, and/or final
123 study reports.

124
125 FDA recognizes that a range of approaches may be used to apply currently supported data
126 standards (e.g., Clinical Data Interchange Standards Consortium’s (CDISC’s) Study Data
127 Tabulation Model (SDTM)) to RWD sources such as EHR or claims data.

128
129 With adequate documentation of the conformance methods used and their rationale, study data
130 derived from RWD can be transformed to SDTM datasets and submitted to FDA in an applicable
131 drug submission.

D. Considerations for Mapping Real-World Data to Study Data Submission Standards

132
133
134
135
136 FDA is aware that, for nearly every *data domain*, there is wide divergence in the terminologies
137 used and their precise meaning between RWD sources and FDA-supported data standards.
138 Examples range from the meaning and specific terms used for race/ethnicity, terminology
139 systems for medications, and interpretation of health care records for vital measurements. Even
140 for seemingly identically recorded variables (e.g., male/female), there can be differences in the
141 way these variables are defined between RWD sources and FDA-supported data standards. For
142 example, sex as a variable may be codified in CDISC’s terminology as a concept based on
143 physical characteristics, whereas EHRs may use gender identity. In such cases, sponsors should
144 document the potential impact of mapping the sex variable or other variables to CDISC’s
145 terminology on the study findings.

146
147 Documentation of the sponsor’s rationale for choosing particular CDISC data elements for RWD
148 and documentation of the differences between the two is critical. The sponsor should provide a
149 description of the general approach and anticipated impact of data mapping as a part of or in an
150 appendix to the *Study Data Reviewer’s Guide* to highlight the domains involved. Furthermore,
151 the sponsor should include a data dictionary that documents the definition of every data element
152 used and all relevant information about the element, such as its relationships to other data, origin,
153 usage, and format. The technical details, best not included in the Study Data Reviewer’s Guide,

¹⁴ The *Study Data Technical Conformance Guide* is available at <https://www.fda.gov/industry/fda-resources-data-standards/study-data-standards-resources>.

Contains Nonbinding Recommendations

Draft — Not for Implementation

154 can be referenced by guiding the reviewers to the detailed mappings in the *Define-XML* file (see
155 the Appendix) and relevant dataset/domains.

156

157 E. Considerations for Data Transformations

158

159 Sponsors may encounter challenges when transforming RWD into data that are consistent with
160 FDA-supported data standards. Examples of these challenges include (but are not limited to)
161 management of semantic concepts (terms) that are present at multiple locations in a health record
162 (such as medication information), inconsistent coding or miscoding of concepts (e.g., drugs or
163 diagnoses), changes in data collection or coding practices (e.g., International Classification of
164 Diseases-9 (ICD-9) and ICD-10 codes) that occurred during the study, or missing information
165 (either because information is not typically recorded in health care settings or due to inconsistent
166 data entry).

167

168 Sponsors should document data challenges encountered during transformation to an FDA-
169 supported data standard and a justification of their approach to enable the application of an FDA-
170 supported data standard. Mapping of standards and terminologies can be handled using the
171 Define-XML (see the Appendix) and domain data files. Given that describing the rationale and
172 justification for approaches used to reconcile any challenges in the source data are likely to
173 require free-text description, in addition, a narrative should be presented in the Study Data
174 Reviewer's Guide, either in the body or as an appendix, with appropriate directions for reviewers
175 to the Define-XML and dataset/domains for more detail, if needed.

176

177

178 IV. GLOSSARY

179

180 **Controlled Terminology:** a finite set of values (e.g., codes, text, numeric) that represent the only
181 allowed values for a data item. Generally, controlled terminology standards specify the key
182 concepts that are represented as definitions, preferred terms, synonyms, codes, and code
183 systems.¹⁵

184

185 **Data Curation:** application of standards (e.g., Clinical Data Interchange Standards Consortium
186 (CDISC), Health Level 7, International Classification of Diseases-10 Clinical Modification
187 (ICD-10-CM)) to source data, for example, the application of codes to adverse events, disease
188 staging, the progression of disease, and other medical and clinical concepts.

189

190 **Data Domain:** a collection of logically related observations (with a common, specific topic) that
191 are normally collected for all subjects in a clinical investigation. NOTE: The logic of the
192 relationship may pertain to the scientific subject matter of the data or to its role in the trial/study.
193 Example domains include laboratory test results, adverse events, concomitant medications.¹⁶

194

195 **Data Standards:** a set of rules about how a particular type of data should be structured, defined,
196 formatted, or exchanged between computer systems. Data standards make submissions

¹⁵ See Glossary at <https://www.cdisc.org/standards/glossary>.

¹⁶ Id.

Contains Nonbinding Recommendations

Draft — Not for Implementation

197 predictable, consistent, and have a form that an information technology system or a scientific
198 tool can use.

199
200 ***Data Transformation:*** includes data extraction, cleansing, and integration (e.g., into a Common
201 Data Model (CDM)).

202
203 ***Define-XML:*** transmits metadata that describes any tabular dataset structure. When used with
204 the CDISC content standards, it provides the metadata for human and animal model datasets
205 using the SDTM and/or Standard for Exchange of Nonclinical Data (SEND) standards and
206 analysis datasets using Analysis Data Model (ADaM).¹⁷

207
208 ***Electronic Health Record (EHR):*** an individual patient record contained within the EHR
209 system. A typical individual EHR may include a patient’s medical history, diagnoses, treatment
210 plans, immunization dates, allergies, radiology images, pharmacy records, and laboratory and
211 test results.

212
213 ***Exchange Format:*** a data format for converting from one file or database structure to another.
214 For example, XML is commonly used as a data exchange format.

215
216 ***Mapping:*** the process of creating data element linkages between two distinct data models.

217
218 ***Medical Claims Data:*** the compilation of information from medical claims that health care
219 providers submit to insurers to receive payment for treatments and other interventions. Medical
220 claims data use standardized medical codes, such as the World Health Organization’s
221 International Classification of Diseases Coding (ICD-CM) diagnosis codes, to identify diagnoses
222 and treatments.¹⁸

223
224 ***National Drug Code (NDC):*** a universal product identifier for drugs in the United States that
225 applies a unique 10-digit or 11-digit, 3-segment number (the first segment identifies the labeler;
226 the second segment is the product code that identifies the specific strength, dosage form and
227 formulation of a drug; and the third segment identifies package sizes and types) to
228 pharmaceuticals.

229
230 ***Non-interventional (observational) study:*** a type of study in which patients are not assigned to a
231 study arm according to a protocol, but instead receive the drug of interest during routine clinical
232 care.

233
234 ***Registries:*** organized systems that collect uniform data (clinical and other) to evaluate specified
235 outcomes for a population defined by a particular disease, condition, or exposure, and that serve
236 one or more scientific, clinical, or policy purposes.

237

¹⁷ See Define-XML at <https://www.cdisc.org/standards/data-exchange/define-xml>.

¹⁸ See *Framework for FDA’s Real-World Evidence Program* (December 2018) at <https://www.fda.gov/media/120060/download>.

Contains Nonbinding Recommendations

Draft — Not for Implementation

238 ***RxNorm:*** provides normalized names for clinical drugs and links its names to many of the drug
239 vocabularies commonly used in pharmacy management and drug interaction software, including
240 those of First Databank, Micromedex, and Gold Standard Drug Database.¹⁹

241
242 ***Source Data:*** all information in original records and certified copies of original records of
243 clinical findings, observations, or other activities in a clinical study necessary for the
244 reconstruction and evaluation of the study. Source data are contained in source documents
245 (original records or certified copies).²⁰

246
247 ***Study Data Reviewer’s Guide:*** a study data reviewer’s guide should describe any special
248 considerations or directions or conformance issues that may facilitate an FDA reviewer’s use of
249 the submitted data and may help the reviewer understand the relationships between the study
250 report and the data.

251
252 ***Terminologies:*** the body of terms used for particular technical application to standardize a
253 medical term for the submission of nonclinical and clinical study data.

254
255 ***Traceability:*** permits an understanding of the relationships between the analysis results (tables,
256 listings, and figures in the study report), analysis datasets, tabulation datasets, and source data.²¹

257

¹⁹ See RxNorm at <https://www.nlm.nih.gov/research/umls/rxnorm/index.html>.

²⁰ See FDA guidance for industry *Use of Electronic Health Record Data in Clinical Investigations* (July 2018).

²¹ See FDA technical specifications document *Study Data Technical Conformance Guide* (June 2021).

Contains Nonbinding Recommendations

Draft — Not for Implementation

APPENDIX: EXAMPLES OF MAPPING HEALTH CARE DATA TO CDISC SDTM

Differences in the coding systems used between real-world data (RWD) and traditional clinical trial data can usually be addressed using the Define-XML file, which is included in all standard Study Data Tabulation Model (SDTM) submissions. The Define-XML file, along with the appropriate use of *Decode* or *Alias* data elements, provides a mechanism for communicating the transformation of external coding systems to the appropriate SDTM controlled terminology.

An example of this approach involves race/ethnicity data, where the Food and Drug administration (FDA) anticipates both heterogeneity among electronic health records (EHRs) as well as between EHR and Clinical Data Interchange Standards Consortium (CDISC) terminologies. In the guidance for industry *Collection of Race and Ethnicity Data in Clinical Trials* (October 2016), FDA recommends that a minimum of five specific categories be used to define race:

- (1) American Indian or Alaska Native
- (2) Asian
- (3) Black or African American
- (4) Native Hawaiian or Other Pacific Islander
- (5) White

RWD sources, however, may not follow the same system of coding. Given that FDA recommends using the race and ethnicity categorization outlined in the October 2016 guidance mentioned above, a sponsor should map the RWD terminology system to the relevant SDTM terminology. To achieve this objective, the *Decode* or *Alias* elements in Define-XML file can be used to document the conversions to a single nomenclature while ensuring *traceability*.

Table 1 illustrates how race can be transformed from non-standardized to standardized data using FDA-supported data standards. In Table 1, the *Decode* column shows the original codes present in an EHR system and the *Code* column shows the relevant mapped term in the current FDA-supported controlled terminology:

Contains Nonbinding Recommendations

Draft — Not for Implementation

303 **Table 1: Approach to Using Define-XML to Indicate Decision Involved in Transforming Non-**
304 **Standardized Data (Race Data) to Standardized Data (i.e., SDTM and ADaM)**

305 *Illustrative example of an approach to representing cross-mapping of coding systems, in this case for Race data, to CDISC*
306 *coding in the Define-XML file. This table does not recommend **how** to map coding systems to CDISC terminology, only how to*
307 *represent the mapping choices made.*
308

| Permitted Value (Code)* | Display Value (Decode)** |
|--|---|
| Race [RACE, C74457] | |
| AMERICAN INDIAN OR ALASKA NATIVE [C41259] | American Indian or Alaska Native, Native American, Native of Alaska |
| ASIAN [C41260] | Asian, Chinese |
| BLACK OR AFRICAN AMERICAN [C16352] | Black or African American, Black |
| NATIVE HAWAIIAN OR OTHER PACIFIC ISLANDER [C41219] | Native Hawaiian or Other Pacific Islander, Samoan |
| OTHER [*] | Other |
| WHITE [C41261] | White, Mexican |

309 *Permitted Value (Code): vocabulary that is provided in the study data tabulations and conformant with controlled terminologies.
310

311 **Display Value (Decode): vocabulary that was used in the original data set (i.e., EHR value). Code/Decode: Respective CDISC
312 elements.
313

314 Differences in **controlled terminology** between RWD systems and FDA submission data
315 standards may make mapping terminology challenging. Furthermore, FDA is aware that in **non-**
316 **interventional studies**, a sponsor may use data aggregated from multiple RWD systems. Such
317 situations can complicate the use of terminologies further, since different RWD sources, such as
318 EHRs, might use different coding systems for the same concept or might use the same coding
319 system but use different default codes for the same item.

320
321 Various approaches can be applied to permit the use of RWD in applicable drug submissions.
322 Examples of potential approaches are: 1) translating the codes to their mapped structured
323 definitions with subsequent mapping to appropriate CDISC controlled terminologies, which
324 provides the most detail but is labor-intensive; or, alternatively 2) mapping all original codes to
325 the least granular analogous codes, and then mapping those to CDISC controlled terminologies,
326 which is less labor-intensive yet necessitates that detail of a more specific categorization will not
327 be represented in the submitted, standardized dataset. It is up to the sponsors to determine the
328 best approach to mediating data transformation, as well as to document and justify their approach
329 accordingly. However, if details that are essential to the consideration of the safety and
330 effectiveness of a drug are absent, the latter approach may not be appropriate. Whatever
331 approach is used, the application of *Decode* to achieve CDISC standard controlled terminology is
332 one mechanism to document the normalization of nomenclature into a format developed by
333 CDISC.
334

335 An example where concepts and terminology do not map precisely and directly is the SDTM
336 intervention domain capturing *drugs prescribed*. Domains containing drugs prescribed data may
337 be mapped to a template SDTM intervention domain. (FDA anticipates that an EHR system may
338 use prescription coding from *RxNorm*, *NDC*, or other such systems.) Additionally, and unless a
339 sponsor uses EHR data where the prescription dispensing information is retained or opts to link
340 EHR information with medical claims data, uncertainty will persist regarding the actual
341 prescription (e.g., whether a generic pharmaceutical agent was substituted). In such cases,

Contains Nonbinding Recommendations

Draft — Not for Implementation

sponsors should apply the *Decode* and/or *Alias* elements within Define-XML to normalize the coding systems (as illustrated in Table 2), where the final dataset (see Table 3) will have prescription dosing information uniformly reflecting the less specific prescription coding system.

The standard modeling would be accompanied by the Define-XML code list as follows:

Table 2: Approach to Using Define-XML to Indicate Decision Involved in Transforming Non-Standardized Data (Dugs Prescribed) to Standardized Data (i.e., SDTM and ADaM)

Illustrative example of an approach to representing cross-mapping of coding systems, in this case for prescription data, to CDISC coding in the Define-XML file. This table does not recommend how to map coding systems to CDISC terminology, only how to represent the mapping choices made.

| Permitted Value (Code)* | Display Value (Decode)** |
|-------------------------|--|
| FLUoxetine | FLUoxetine 40 milligram (mg) Oral Capsule, RxCUI = 383919, 0093-7198-56, FLUoxetine 40 mg Oral Capsule Generic Permitted |
| PROzac | PROzac 40 mg Oral Capsule, RxCUI = 313989, 0777-3107-30 |

*Permitted Value (Code): vocabulary that is provided in the study data tabulation.

**Display Value (Decode): vocabulary that was used in the original data set (i.e., EHR value). Code/Decode: Respective CDISC elements.

The standard modeling of prescription data in the domain file would appear as shown below (Table 3):

Contains Nonbinding Recommendations

Draft — Not for Implementation

385 **Table 3: Example of the Drugs Prescribed Data in the Respective Domain Upon Mapping**
 386 *Illustrative example of an approach to representing the values from data, in this case for prescription data, to CDISC format in*
 387 *the relevant data domain file. This table does not recommend **how** to map data to CDISC standards, only how to represent the*
 388 *mapping choices made. The column headers represent CDISC data elements. For more information, see*
 389 *<https://www.cdisc.org/standards/foundational/sdtm>.*
 390

| Original Value (from original dataset) | --TRT | --MODIFY | --DECOD | --DOSE | --DOSU | --ROUTE | --DOSFRM |
|---|------------|--|--|--------|--------|---------|----------|
| FLUoxetine 40 mg Oral Capsule | FLUoxetine | To be populated by the sponsor to assist in coding to standard terminology | Generic Drug Name in WHO Drug <i>(either from original data system or assigned by medical coding vendor for sponsor)</i> | 40 | mg | ORAL | CAPSULE |
| FLUoxetine 40 mg Oral Capsule Generic Permitted | FLUoxetine | (same as above) | (same as above) | 40 | mg | ORAL | CAPSULE |
| RxCUI = 383919 | FLUoxetine | (same as above) | (same as above) | | | | |
| 0093-7198- 56 | FLUoxetine | (same as above) | (same as above) | | | | |
| PROzac 40 mg Oral Capsule | PROzac | (same as above) | (same as above) | 40 | mg | ORAL | CAPSULE |
| RxCUI = 313989 | PROzac | (same as above) | (same as above) | | | | |
| 0777-3107- 30 | PROzac | (same as above) | (same as above) | | | | |

391
 392 Although only a few examples are presented here, sponsors should use elements of the Define-
 393 XML file and relevant domain data files to communicate how the health care terminology of all
 394 **data domains** were normalized to CDISC standard terminology. As in the examples shown
 395 above, the technical details of all transformations are best placed in the data files.