# Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products

## 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***DRAFT GUIDANCE***

**This guidance document is being distributed for comment purposes only.**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be submitted within 90 days of publication in the *Federal Register* of the notice announcing the availability of the draft guidance.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, contact (CDER) Ansalan Stewart, 240-402-6631, or (CBER) Office of Communication, Outreach and Development, 800-835-4709 or 240-402-8010.

**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)**
**Oncology Center of Excellence (OCE)**

**November 2021**
**Real World Data/Real World Evidence (RWD/RWE)**

# Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products

## 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)**
**Oncology Center of Excellence (OCE)**

**November 2021**
**Real World Data/Real World Evidence (RWD/RWE)**

# TABLE OF CONTENTS

# Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry[1]

This draft guidance, when finalized, will represent the current thinking of the Food and Drug Administration (FDA or Agency) on this topic.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

## I. INTRODUCTION

The 21st Century Cures Act (Cures Act),[2] signed into law on December 13, 2016, was intended to accelerate medical product development and bring innovations and advances faster and more efficiently to the patients who need them. Among other provisions, the Cures Act added section 505F to the Federal Food, Drug, and Cosmetic Act (FD&C Act). In response to the requirements in section 505F, the Food and Drug Administration (FDA) created a framework for a Real-World Evidence (RWE) Program to evaluate the potential use of real-world evidence to help support the approval of a new indication for a drug[3] already approved under section 505(c) of the FD&C Act or to help support or satisfy postapproval study requirements.[4]

This guidance provides sponsors and other stakeholders with considerations when either proposing to design a registry or using an existing registry to support regulatory decision-making about a drug's effectiveness or safety. This guidance does not provide recommendations on choice of study design or type of statistical analysis when analyzing data from registries (*registry data*).

FDA is issuing this guidance as part of its RWE Program and to satisfy, in part, the mandate under section 505F of the FD&C Act to issue guidance on the use of RWE in regulatory

[1] This guidance has been prepared by the Office of Medical Policy in the Center for Drug Evaluation and Research in cooperation with the Center for Biologics Evaluation and Research and the Oncology Center of Excellence 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 the *Framework for FDA's Real-World Evidence Program*, available at https://www.fda.gov/media/120060/download. In addition to drug and biological products approved under section 505(c) of the FD&C Act, this framework is also intended for application to biological products licensed under the Public Health Service Act.

decision-making.[5] For the purposes of this guidance, FDA defines real-world data (RWD) and RWE as follows:

- RWD are data relating to patient health status and/or the delivery of health care routinely collected from a variety of sources.

- RWE is the clinical evidence about the usage and the potential benefits or risks of a medical product derived from analysis of RWD.

Topics covered in this guidance include:

- Considerations regarding a registry's fitness-for-use in regulatory decision-making, focusing on attributes of a registry that support the collection of relevant and reliable data

- Considerations when linking a registry to another data source for supplemental information, such as data from medical claims, electronic health records (EHRs), digital health technologies, or other registries

- Considerations for supporting FDA review of submissions that include registry data

Whether registry data are fit-for-use in regulatory decision-making depends on the attributes that support the collection of relevant and reliable data (described in this guidance) as well as additional scientific considerations related to study design and study conduct that are beyond the scope of this guidance.

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

## II. BACKGROUND

For the purposes of this guidance, a registry is defined as an organized system that collects clinical and other data in a standardized format for a population defined by a particular disease, condition, or exposure.[6] Establishing registries involves enrolling a predefined population and collecting prespecified health-related data for each patient in that population (patient-level data).

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

[6] Registries can generally be categorized as (1) disease registries that use the state of a particular disease or condition as the inclusion criterion, (2) health services registries where the patient is exposed to a specific health care service, or (3) product registries where the patient is exposed to a specific health care product.

Data about this population can be entered directly into the registry (e.g., clinician-reported outcomes) and can also include additional data linked from other sources that characterize registry participants. Such external data sources can include data from medical claims, from pharmacy and/or laboratory databases, and from EHRs, blood banks, and/or medical device outputs. Trained staff should follow standard operating procedures to aggregate data for a registry and carry out **data curation**.[7]

Registries range in complexity regarding the extent and detail of the data captured and how the data are curated. For example, registries used for quality assurance purposes related to the delivery of care for a particular health care institution or health care system tend to collect limited data related to the provision of care. Registries designed to address specific research questions tend to systematically collect longitudinal data in a defined population, on factors characterizing patients' clinical status, treatments received, and subsequent clinical events. The data collected in a given registry and the procedures for data collection are relevant when considering how registry data can be used.

Registries have the potential to support medical product development, and registry data can ultimately be used, when appropriate, to inform the design and support the conduct of either interventional studies (clinical trials) or non-interventional (observational) studies. Examples of such uses include, but are not limited to:

- Characterizing the natural history of a disease[8]

- Providing information that can help determine sample size, selection criteria, and study **endpoints** when planning an interventional study

- Selecting suitable study participants—based on factors such as demographic characteristics, disease duration or severity, and past history or response to prior therapy—to include in an interventional study (e.g., randomized trial) that will assign a drug to assess that drug's safety or effectiveness

- Identifying biomarkers or clinical characteristics that are associated with important clinical outcomes of relevance to the planning of interventional and non-interventional studies

- Supporting, in appropriate clinical circumstances, inferences about safety and effectiveness in the context of:

[7] Words and phrases in **bold text** are defined in the Glossary.

[8] For the purposes of this guidance, a *natural history study* is a non-interventional (observational) study intended to track the course of the disease for purposes such as identifying demographic, genetic, environmental, and other (e.g., treatment) variables that correlate with disease development and outcomes. Natural history studies are likely to include patients receiving the current standard of care and/or emergent care, which may alter some manifestations of the disease. Disease registries are common platforms to acquire the data for natural history studies.

- A non-interventional study evaluating a drug received during routine medical practice and captured by the registry

- An externally controlled trial including registry data as an external control arm[9]

An existing registry can be used to collect data for purposes other than those originally intended, and reusing a registry's infrastructure to support multiple interventional and non-interventional studies can generate efficiencies. Before designing and initiating an interventional or non-interventional study using registry data for regulatory decisions, sponsors should consult with the appropriate FDA review division regarding the appropriateness of using a specific registry as a real-world data source.

## III. DISCUSSION

### A. Using Registry Data to Support Regulatory Decisions

Registry data can have varying degrees of suitability within a regulatory context, depending on several factors, including how the data are intended to be used for regulatory purposes; the patient population enrolled; the data collected; and how registry datasets are created, maintained, curated, and analyzed. Registry data collected initially for one purpose (e.g., to obtain comprehensive clinical information on patients with a particular disease) may or may not be fit-for-use for another purpose (e.g., to examine a drug-outcome association in a subset of these patients).

Sponsors should consider both the strength and limitations of using registries as a source of data to generate evidence for regulatory decision-making. Registries may have advantages over other RWD sources, given that registries collect structured and predetermined **data elements** and can offer longitudinal, curated data about a defined population of patients and their corresponding disease course, complications, and medical care. In addition, registries can systematically collect patient-reported data that medical claims datasets or EHR datasets may lack.

Registries can have limitations for use in a regulatory context. For example, existing registries may focus on one disease, with limited information on comorbid conditions, even after linkage to other data sources. In addition, the enrolled patients may not be representative of the target population of interest due to challenges related to patient recruitment and retention. For example, patients with more severe disease may be more likely to be enrolled in a registry compared to patients with milder disease; or enrolled patients might have different self-care practices, socioeconomic backgrounds, or levels of supportive care versus the entire population of interest. These issues can potentially introduce bias into analyses that make use of registry

[9] An *externally controlled trial*, as one type of clinical trial, compares outcomes in a group of participants receiving the test treatment with outcomes in a group external to the trial, rather than to an internal control group from the same trial population assigned to a different treatment. The external control arm can be a group, treated or untreated, from an earlier time (historical control) or a group, treated or untreated, during the same time period (concurrent control) but in another setting.

data. Additional potential limitations of registries involve issues with data heterogeneity (e.g., different clinical characteristics across various populations) and variation in approaches used to address data quality.

In general, registries are better suited as a data source for regulatory purposes when sponsors aim to capture objective endpoints, such as death or hospitalization. Subjective endpoints, such as pain, can be collected in a registry, but additional challenges are involved to standardize such measurements. In addition, a registry that is designed to collect data to answer a specific research question can have advantages over an existing registry designed for another purpose, which is subsequently repurposed for that same question. A key advantage of a registry developed to answer a specific research question is that developers of such a registry can anticipate collecting specific information about clinical endpoints and outcomes, whereas an existing registry may need to be linked to other data sources.

Before using any RWD (including registry data) for regulatory decision-making, sponsors should consider whether the data are fit-for-use by assessing the data's relevance and reliability. For the purposes of this guidance, the term *relevance* includes the availability of key data elements (patient characteristics, exposures, outcomes) and a sufficient number of representative patients for the study, and the term *reliability* includes **data accuracy, completeness, provenance**, and **traceability**.

### B. Relevance of Registry Data

When considering whether to use an existing registry [10] for regulatory purposes, a sponsor's overall assessment of the relevance of registry data should consider whether the registry is adequate for evaluating the scientific objectives. As a part of this assessment, sponsors should carefully consider the data elements captured by the registry.

The specific data elements that should be captured by a registry depend on the sponsor's intended use or uses of the registry. For example, the minimum set of data elements in a registry may need to be more comprehensive if the sponsor intends to use the registry data for an external control arm in an externally controlled trial, compared to if the sponsor intends to use the registry to enroll participants in an interventional study. The registry should retain information documenting any data elements that are no longer being collected or new data elements that begin to be collected. Sponsors also should develop a plan to reduce loss to follow-up of registry participants.

The assessment of the data's relevance is context dependent. For example, when considering using a registry for regulatory purposes, sponsors should consider the methods involved in patient selection and the effect those methods have on the representativeness of the population in the registry. In particular, the inclusion and exclusion criteria used to enter patients into a registry may result in the patient population in a registry study differing from the target population for the sponsor's drug development program. Furthermore, patients who remain

[10] An existing registry can be used as is or modified for specific research purposes, such as by adding a module to capture an outcome of interest for longer follow-up.

enrolled in the registry may differ from those who do not remain (e.g., having experienced different adverse events).

Registries generally include data elements that capture information about patient characteristics, treatments received, and health outcomes for patients enrolled in the registry. Such information typically includes a unique patient identifier; the date of patient consent to participate in the registry; and baseline characteristics of the patient at that time, such as demographic factors, comorbidities, medical history, and other information. Sponsors should consider which data elements a registry should have based on their intended use of the registry.

The following are non-exhaustive examples of potential data to include in a registry:

- Demographic and clinical information:
  - Patient demographic factors, including date of birth, gender, race and ethnicity, height, weight, smoking status, alcohol use, and recreational drug use
  - Primary diagnosis of interest, including date of diagnosis, test name and result, diagnostic code, and genetic or other testing if available; specific approach to capture grade, severity, and/or burden of disease and important milestones in disease progression
  - Patient comorbidities, including current status (e.g., complications, disease manifestations) of those diseases, dates of assessments, and therapies for individual comorbid conditions
  - Additional relevant information regarding characteristics thought to modify disease severity or progression
- Treatment information for the disease of interest (as applicable):
  - Chemical name and product name of the drug or drugs
  - Formulation and dosage, start and end dates of each treatment, and reason for discontinuation (as applicable)
  - Type and date of procedure or procedures periprocedural complications
- Health-related outcome information:
  - Specific clinical events (e.g., heart attack, stroke, hospitalization, death) of interest and date of occurrence
  - Other clinical outcomes (e.g., disease progression, relapse, disability, functional status, quality of life measure) and date of occurrence

241 – Changes in patient management and date of occurrence
242
243 • Pregnancy-related information,[11] when intending to collect data related to pregnancy or
244 pregnancy outcomes:
245
246 – Prior pregnancy history
247
248 – Date of last menstrual period, if known, and ultrasound reports that assess gestational
249 age
250
251 – Gestational timing of drug exposure
252
253 – Maternal outcomes, including pre-eclampsia, eclampsia, etc.
254
255 – Pregnancy outcomes, including live birth, stillbirth, miscarriage, etc.
256
257 – Fetal outcomes, including major congenital malformations, small-for-gestational age,
258 preterm birth, low birth weight, any other relevant adverse fetal outcomes, etc.
259
260 **C. Reliability of Registry Data**
261
262 When considering using an existing registry or establishing a registry de novo, sponsors should
263 ensure there are processes and procedures to govern registry operation, education and training of
264 registry staff, resource planning, and general practices that help ensure the quality of the registry
265 data. Such governance attributes help ensure that the registry can achieve its objectives and
266 should include, but not be limited to:
267
268 • An established data dictionary and rules for the **validation** of queries and edit checks of
269 registry data (as applicable), to be made available for those who intend to use the registry
270 data to perform analyses
271
272 • Defined processes and procedures for the registry, such as:
273
274 – Data collection, curation, management, and storage, including processes in place to
275 help ensure that data within a registry can be confirmed by source data (as applicable)
276 for that registry
277
278 – Plans for how patients, researchers, and clinicians will access and interact with the
279 registry data and the registry's data collection systems

[11] For further discussion of the design of pregnancy safety studies, including recommended data elements, see the draft guidance for industry *Postapproval Pregnancy Safety Studies* (May 2019). For further discussion of clinical lactation studies, see the draft guidance for industry *Clinical Lactation Studies: Considerations for Study Design* (May 2019). When final, these guidances will represent FDA's current thinking on these topics. 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.

  - Terms and conditions for use of the registry data by parties other than the registry creator (e.g., terms and conditions a sponsor should satisfy to permit combining the registry data with data from another source)

- Conformance with 21 CFR part 11, as applicable, including maintenance of access controls and **audit trails** to demonstrate the provenance of the registry data and to support traceability of the data[12]

Sponsors also should ensure that a registry adheres to applicable jurisdictional human subject protection requirements, including protecting the privacy of patient health information, when designing a registry and developing protocols for the subsequent use of the data from the registry. FDA also recommends that an institutional review board or independent ethics committee be consulted when developing a registry to review data collection and other procedures associated with the registry.

Factors that FDA considers when assessing the reliability of registry data include how the data were collected (**data accrual**). FDA also considers whether the registry personnel and processes in place during data collection and analysis provide adequate assurance that errors are minimized and that **data integrity** is sufficient. Sponsors should address whether the registry has privacy and security controls in place to ensure that the confidentiality and security of data are preserved.[13] When sponsors intend to capture patient-reported outcomes (PROs) in a registry, sponsors should review the recommendations in FDA's guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Changes* (December 2009).

To support the collection of reliable data within a registry, a registry's data dictionary should include:

- Data elements and how the data elements are defined
- Ranges and allowable values for the data elements
- Reference to the source data for the data elements

Sponsors are encouraged to use **common data elements** to promote standardized, consistent, and universal data collection. Such an approach can facilitate comparing or linking registry data to

[12] For additional discussion on the use of electronic records and electronic signatures under part 11, see the draft guidance for industry *Use of Electronic Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 — Questions and Answers* (June 2017). When final, this guidance will represent FDA's current thinking on this topic.

[13] For recommendations on controls to ensure confidence in the reliability, quality, and integrity of electronic source data in FDA-regulated clinical investigations, see the guidance for industry *Electronic Source Data in Clinical Investigations* (September 2013).

data from other sources. When possible, such standardized terminology and the associated **data standards** used by the registry should be consistent with those used by FDA.[14]

Appropriate policies and procedures should be in place to support the reliability of the registry data, including prespecifying data validation rules for queries and edit checks of registry data, as well as validating the electronic systems used to collect registry data.[15] Additional policies and procedures should be in place that enable FDA and persons interested in using the registry's data to assess the quality of the data, including to help address issues such as errors in coding or interpretation of the source document or documents, as well as data entry, transfer, or **transformation** errors. The formats and definitions of the data entered in the registry should be consistent over time, and any changes in diagnostic criteria or clinical definitions over time should be accounted for and documented.

Registries in the form of an electronic database should have safeguards in place, including data management strategies, to support data assurance. Data management strategies should include processes and procedures to:

- Implement and maintain version control by documenting the date, time, and originator[16] of data entered in the registry; performing preventative and/or corrective actions to address changes to the data (including flagging erroneous data without deleting the erroneous data, while inserting the corrected data for subsequent use); and describing reasons for any changes to data without obscuring previous entries.

- Ensure data transferred from another data format or system are not altered in the migration process

- Seek to integrate data in the registry that were previously collected using data formats or technology (e.g., operating systems, hardware, software) that are now outdated

- Account for changes in clinical information over time (e.g., criteria for disease diagnosis, cancer staging)

- Explain the auditing rules and methods used and the mitigation strategies used to reduce errors

[14] FDA has specific data standards (describing a standard way to exchange clinical study data) and terminology recommendations for marketing applications. See FDA's Study Data Standards Resources web page, available at https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm.

[15] Validation of electronic systems may include, but is not limited to, demonstrating correct installation of the electronic system and testing of the system to ensure that it functions in the manner intended. This topic is also discussed in the draft guidance for industry *Use of Electronic Records and Electronic Signatures in Clinical Investigations Under 21 CFR Part 11 – Questions and Answers*. When final, this guidance will represent FDA's current thinking on this topic.

[16] Source data originators include persons, systems, devices, and instruments. For additional information, see the guidance for industry *Electronic Source Data in Clinical Investigations* (September 2013).

- Describe the types of errors that were identified based on audit findings and how the data were corrected

Indicators of **data consistency**, accuracy, and completeness should be assessed periodically, with the frequency dependent on the purposes of the registry data (e.g., for the sole purpose of facilitating recruitment in a randomized controlled trial versus using the registry data in an interventional or non-interventional study analysis). Routine descriptive statistical analyses should be performed to detect the extent of any missing data, inconsistent data, outliers, and losses to follow-up.

### D. Considerations When Linking a Registry to Another Registry or Another Data System

When a registry does not capture all the necessary information to answer the question of interest in an interventional or non-interventional study, sponsors may consider obtaining supplemental information from another source. For example, sponsors may consider linking the data on a patient in the registry to the same patient in another data system or systems, such as another registry, an EHR, a medical claims database, or through **digital health technologies** (DHTs), such as sensors that allow for continuous or intermittent recording of physiological and/or behavioral data (e.g., blood pressure, physical activity, glucose levels) or software applications that are run on general-purpose computing platforms.

If a registry is to be populated with data from another data system, sponsors should consider the potential impact of the additional data on overall integrity of the registry data. Sponsors should use strategies to correct for redundant data, to resolve any inconsistencies in the data, and to address other potential problems, such as the ability to protect patient privacy while transferring data securely. Sponsors should have a plan for addressing the adequacy of patient-level linkages (i.e., that the same patient is being matched). Sponsors also should consider any jurisdictional requirements (e.g., country-specific laws) when seeking to link patient-level data to another registry or data system.

Sponsors should also consider whether the data sources to be linked are interoperable and support appropriate informatics strategies to ensure data integration. Sponsors should ensure that (1) sufficient testing is conducted to demonstrate interoperability of the linked data systems, (2) the automated electronic transmission of data elements to the registry functions in a consistent and repeatable fashion, and (3) data are accurately, consistently, and completely transmitted. Predefined rules to check for logical consistency and value ranges should be used to confirm that data within a registry were retrieved accurately from a linked data source and that the operational definitions for the linked variables are aligned.

Documentation of the process sponsors used to validate the transfer of data should be available for FDA to review during sponsor inspections. Sponsors should also ensure that software updates to the registry database or additional data sources do not affect the integrity,

interoperability, and security of data transmitted to the registry.[17] For example, issues such as the correct temporal alignment of linked data and registry data should be considered.

The appropriateness of using additional data sources also depends on how the sponsor intends to use the linked data and the ability to obtain similar data for all patients. For example, for each potential data source, the sponsor should consider whether:

- The linkage is appropriate for the proposed research question (e.g., the additional data source provides relevant clinical detail and/or long-term follow-up information)
- The data can be accurately matched to patients in the registry and whether linking records between the two (or more) databases can be performed accurately
- The variables of interest in the registry and additional data sources have consistent definitions and reliable ascertainment approaches
- The data have been captured with sufficient accuracy, consistency, and completeness to meet registry objectives

After a sponsor decides to use an additional data source or sources to supplement the registry, the sponsor should develop the approach and algorithms needed to link such data to a registry. Additionally, the sponsor should determine how data integrity will be evaluated, including how assessments of any inaccuracies introduced by the linkage (e.g., overcounts of a particular data measure) will be made. The sponsor also should use appropriate methods for data entry, coding, cleaning, and transformation for each linked data source.

### E. Considerations for Regulatory Review

Sponsors interested in using a specific registry as a data source to support a regulatory decision should meet with the relevant FDA review division before conducting a study that will include registry data.[18] Sponsors should confer with FDA regarding (1) the ability to accurately define and evaluate the target population based on the planned inclusion and exclusion criteria; (2) which data elements will come from the registry (versus other data sources) and their adequacy, as well as the frequency and timing of data collection; (3) the planned approach for linking the registry to another registry or other data system, when linking is anticipated; (4) the planned methods to ascertain and validate outcomes, including diagnostic requirements and the level of validation or adjudication of outcomes FDA agrees is needed; and (5) the planned methods to validate the diagnosis of the disease being studied.

[17] See footnote 15.

[18] For example, sponsors can request a Type C meeting for non-interventional studies. FDA issued the draft guidance for industry regarding formal meetings with FDA, *Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products* (December 2017). When final, this guidance will represent FDA's current thinking on this topic.

Sponsors should submit protocols and statistical analysis plans for FDA review and comment before conducting an interventional or a non-interventional study when including data from registries. All essential elements of a registry study's design, analysis, and conduct should be predefined, and for each study element, the protocol should describe how that element will be ascertained from the selected RWD source or sources.

Sponsors seeking to use registry data to support a product's effectiveness and safety in a marketing application should ensure that patient-level data are provided to FDA in accordance with applicable legal and regulatory requirements.[19] If the registry data are owned and controlled by third parties, sponsors should have agreements in place with those parties to ensure that all relevant patient-level data can be provided to FDA and that source records necessary to verify the RWD are made available for inspection as applicable.[20]

[19] See, e.g., 21 U.S.C. 355, 42 U.S.C. 262, and 21 CFR 314.50 and 601.2.

[20] See 21 CFR 312.58.

# GLOSSARY

The following terms are defined for the purposes of this guidance document as follows:

**Accuracy**: Correctness of collection, transmission, and processing of data.

**Audit Trail**: A process that captures details of information, such as additions, deletions, or alterations, in an electronic record without obscuring the original record. An audit trail facilitates the reconstruction of the course of such details relating to the electronic record. [21]

**Common Data Elements**: Discrete, clearly defined, and reusable data collection units.[22]

**Data Accrual**: The process by which the data was collected.

**Data Completeness**: The presence of the necessary data to address the study question, design, and analysis.[23]

**Data Consistency**: Relevant uniformity in data across clinical sites, facilities, departments, units within a facility, providers, or other assessors.[24]

**Data Curation**: Application of standards (e.g., Clinical Data Interchange Standards Consortium (CDISC), Health Level 7, ICD-10-CM) to source data; for example, the application of codes to adverse events, disease staging, the progression of disease, and other medical and clinical concepts in an EHR.

**Data Element**: A piece of data corresponding to one patient within a data field.[25]

[21] Guidance for industry *Electronic Source Data in Clinical Investigations* (September 2013).

[22] Kush, RD et al., 2020, FAIR Data Sharing: The Roles of Common Data Elements and Harmonization, J Biomed Inform, 107(1–10), https://doi.org/10.1016/j.jbi.2020.103421.

[23] National Institutes of Health Collaboratory, 2014, Assessing Data Quality for Healthcare Systems Data Used in Clinical Research, accessed November 24, 2021, https://dcricollab.dcri.duke.edu/sites/NIHKR/KR/Assessing-data-quality_V1%200.pdf#search=Assessing%20data%20quality.

[24] Ibid.

[25] Daniel, G, C Silcox, J Bryan, M McClellan, M Romine, and K Frank, 2018, Characterizing RWD Quality and Relevancy for Regulatory Purposes, Duke Margolis Center for Health Policy, accessed November 24, 2021, https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf.

**Data Integrity**: The completeness, consistency, and accuracy of data.[26]

**Data Standard**: A set of rules on how a particular type of data should be structured, defined, formatted, or exchanged between computer systems.[27]

**Data Transformation**: Includes data extraction, cleansing, and integration.

**Digital Health Technology** (**DHT**): A system that uses computing platforms, connectivity, software, and sensors for health care and related uses. These technologies span a wide range of uses, from applications in general wellness to applications as a medical device. They include technologies intended for use as a medical product, in a medical product, or as an adjunct to other medical products (devices, drugs, and biologics). They may also be used to develop or study medical products.[28]

**Endpoint**: A precisely defined variable intended to reflect an outcome of interest that is statistically analyzed to address a particular research question. A precise definition of an endpoint typically specifies the type of assessments made, the timing of those assessments, the tools used, and possibly other details, as applicable.[29]

**Provenance**: An audit trail that "accounts for the origin of a piece of data (in a database, document or repository) together with an explanation of how and why it got to the present place."[30]

**Traceability**: Permits an understanding of the relationships between the analysis results (tables, listings, and figures in the study report), analysis datasets, tabulation datasets, and source data.[31]

**Validation**: The process of establishing that a method is sound or that data are correctly measured.[32]

---

[26] Guidance for industry *Data Integrity and Compliance with Drug CGMP: Questions and Answers* (December 2018).

[27] CDER Data Standards Program, accessed January 13, 2021, https://www.fda.gov/drugs/electronic-regulatory-submission-and-review/cder-data-standards-program.

[28] BEST (Biomarkers, EndpointS, and other Tools) Resource Glossary, 2016, accessed October 14, 2021, available at https://www.ncbi.nlm.nih.gov/books/NBK326791/.

[29] Ibid.

[30] Gupta, A, 2009, Encyclopedia of Database Systems, available at https://doi.org/10.1007/978-0-387-39940-9_1305.

[31] FDA technical specifications document, October 2019, *Study Data Technical Conformance Guide*, available at https://www.fda.gov/media/131872/download,

[32] Porta, M, editor, A Dictionary of Epidemiology, 6th Edition, New York: Oxford University Press, 2014