

---

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

**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)**

**December 2023  
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., 4<sup>th</sup> Floor  
Silver Spring, MD 20993-0002  
Phone: 855-543-3784 or 301-796-3400; Fax: 301-431-6353  
Email: [druginfo@fda.hhs.gov](mailto: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](mailto: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)**

**December 2023  
Real World Data/Real World Evidence (RWD/RWE)**

*Contains Nonbinding Recommendations*

**TABLE OF CONTENTS**

<b>I.</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>II.</b>	<b>BACKGROUND .....</b>	<b>2</b>
<b>III.</b>	<b>DISCUSSION .....</b>	<b>4</b>
	<b>A. Using Registry Data to Support Regulatory Decisions.....</b>	<b>4</b>
	<b>B. Relevance of Registry Data .....</b>	<b>5</b>
	<b>C. Reliability of Registry Data.....</b>	<b>7</b>
	<b>D. Considerations When Linking a Registry to Another Registry or Another Data System ....</b>	<b>10</b>
	<b>E. Considerations for Regulatory Review .....</b>	<b>12</b>

# **Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products Guidance for Industry<sup>1</sup>**

This guidance represents 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 office responsible for this guidance as listed on the title page.

## **I. INTRODUCTION**

The 21st Century Cures Act (Cures Act),<sup>2</sup> 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, FDA created a framework for a Real-World Evidence (RWE) Program to evaluate the potential use of RWE to help support the approval of a new indication for a drug<sup>3</sup> already approved under section 505(c) of the FD&C Act or to help support or satisfy postapproval study requirements.<sup>4</sup> In the context of this program, this guidance provides considerations for sponsors proposing to design a new registry or use 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 statistical methods used to analyze data from registries.

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 decision-making.<sup>5</sup> FDA defines real-world data (RWD) and RWE as follows:

---

<sup>1</sup> This guidance has been prepared by 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.

<sup>2</sup> Public Law 114-255.

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

<sup>4</sup> 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.

<sup>5</sup> See section 505F(e) of the FD&C Act.

## *Contains Nonbinding Recommendations*

- 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),<sup>6</sup> digital health technologies,<sup>7</sup> 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, in part, on attributes that support the collection of relevant and reliable data (described in this guidance).

In general, FDA's guidance documents do not establish legally enforceable responsibilities. Instead, guidances describe the Agency's current thinking on a topic and should be viewed only as recommendations, unless specific regulatory or statutory requirements are cited. The use of the word *should* in Agency guidances 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 drug exposure. Establishing registries involves enrolling a predefined population and collecting prespecified health-related data for each patient in that population (patient-level data). Data about this population can be entered directly into the registry (e.g., clinician- or patient-reported) and can also include data from other sources that characterize registry

---

<sup>6</sup> For additional discussion, see the draft guidance for industry *Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products* (September 2021). When final, this guidance will represent FDA's current thinking on this topic. 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>.

<sup>7</sup> For additional discussion on the use of digital health technologies in clinical investigations, see the draft guidance for industry *Digital Health Technologies for Remote Data Acquisition in Clinical Investigations* (December 2021). When final, this guidance will represent FDA's current thinking on this topic.

## *Contains Nonbinding Recommendations*

participants. Such external data sources can include, but are not limited to, data from medical claims, laboratory or vital statistics databases, EHRs, or medical device outputs.

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 and the standard operating procedures for data collection and management are relevant when considering how registry data can be used.

Registries have the potential to support sponsors' drug development programs, and registry data can be used 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<sup>8</sup>
- 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)
- Identifying biomarkers or clinical characteristics that are associated with clinical outcomes of relevance to support the planning of interventional and non-interventional studies
- Supporting inferences about safety and effectiveness when:
  - Ascertaining clinical outcomes or other clinical data in an interventional study
  - Including registry data as an external control for an interventional trial<sup>9</sup>

---

<sup>8</sup> Disease registries are common platforms to acquire the data for natural history studies intended to understand aspects of the disease such as prognostic variables 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.

<sup>9</sup> For additional discussion on the use of external controls, see the draft guidance for industry *Considerations for the Design and Conduct of Externally Controlled Trials for Drug and Biological Products* (February 2023). When final, this guidance will represent FDA's current thinking on this topic.

## *Contains Nonbinding Recommendations*

- Evaluating a drug received during routine medical practice (e.g., to evaluate clinical outcomes in populations underrepresented in clinical trials)

### **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 regulatory intent (e.g., to evaluate a specific safety risk, to provide information on long-term efficacy outcomes, or to serve as an external control for an interventional study to inform effectiveness); the patient population enrolled; the data collected; and how registry datasets were created, maintained, curated, and linked to other datasets. 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 proposing to use registry data to support regulatory decision-making by FDA are responsible for ensuring that attributes of the registry support the collection of relevant and reliable data, including in situations where the data are from a registry not managed or designed by the sponsor.

An existing registry can be used for purposes other than those originally intended, and reuse of a registry's infrastructure to support multiple interventional and non-interventional studies can generate efficiencies, such as when the number of eligible patients is limited. Sponsors proposing to use data from an existing registry should ensure that the informed consent provided by patients participating in the registry allows for the sponsors' proposed use of the registry data or that patients are re-consented for the proposed use.

A registry that is designed prospectively to collect data to answer a specific research question can have advantages over an existing registry that is repurposed to answer the same question, because developers of a registry intended to answer a specific research question can design the registry to collect drug/environmental exposures, outcomes, and covariates tailored to the research question of interest.

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 predefined 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 data that medical claims datasets or EHR datasets may lack (e.g., patient-reported outcomes, treatment adherence, measures of disease severity).

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

## ***Contains Nonbinding Recommendations***

example, patients with more severe disease may be more likely to be enrolled in a registry compared to patients with milder disease or vice versa, depending on the requirements for registry participation; or enrolled patients might have different self-care practices, socioeconomic backgrounds, or levels of supportive care compared with the entire population of interest. These issues, if not addressed, can potentially introduce bias into analyses that make use of registry data.

As with other sources of RWD, registries are generally better suited as a data source for regulatory purposes when sponsors aim to capture objective data on clinical events such as death or hospitalization. Subjective data, such as pain scores, can be collected in a registry, but additional challenges are involved to standardize such measurements.<sup>10</sup> In addition, for both objective and subjective data, sponsors should consider whether registry participants' knowledge of treatment received could introduce bias into the study findings.

Before using registry data in a study intended to support regulatory decision-making by FDA, 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 data for key study variables (exposures, outcomes, covariates) and sufficient numbers of representative patients for the study, and the term *reliability* includes accuracy, completeness, and traceability.<sup>11</sup>

Sponsors should consult with the appropriate FDA review division regarding the appropriateness of using a specific registry as an RWD source for studies that are intended to provide evidence in support of a regulatory decision by FDA. Sponsors should also submit the protocols and statistical analysis plans to the Agency before conducting the study.

### **B. Relevance of Registry Data**

When considering whether to use an existing registry for regulatory purposes, a sponsor's overall assessment of the relevance of the registry data should ensure that the registry is adequate for evaluating the scientific objectives. As a part of this assessment, sponsors should evaluate 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

---

<sup>10</sup> For additional discussion on how FDA reviews and evaluates patient-reported outcome instruments used to support claims in approved medical product labeling, see the guidance for industry *Patient-Reported Outcome Measures: Use in Medical Product Development to Support Labeling Claims* (December 2009). This guidance will be replaced by a series of guidances on patient-focused drug development (PFDD) when the PFDD guidances are finalized. For additional information, see <https://www.fda.gov/drugs/development-approval-process-drugs/fda-patient-focused-drug-development-guidance-series-enhancing-incorporation-patients-voice-medical>.

<sup>11</sup> Traceability is the method (e.g., audit trail) that allows for knowledge of data provenance (i.e., the origin of a piece of data and how it got into the registry).



## *Contains Nonbinding Recommendations*

to identify participants for an interventional study. Sponsors should confirm that the registry supports data consistency by capturing data elements uniformly, using the same formats and definitions, for a defined patient population and a specific time period. The registry should document any data elements that are no longer being collected, new data elements that have begun to be collected, and any changes in formats or definitions over time.

The assessment of the data's relevance is context dependent. When assessing the relevance of a registry for a specific regulatory purpose, sponsors should consider the methods used to identify and retain registry patients 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 differing from the target population for the sponsor's drug development program.

A registry should have a plan to reduce missing assessments and to minimize loss to follow-up of registry participants. For example, a registry should have plans for outreach to patients who do not maintain contact with the registry or have difficulty doing so (e.g., text messages, phone calls, emails, offers of transportation assistance). Sponsors should consider factors that may influence which patients remain enrolled in the registry and how they may differ from those who do not remain (e.g., because of factors such as adverse events experienced, disease progression, or location of care). Sponsors should consider the extent and reasons for missing assessments and for loss to follow-up to understand the potential impact of missing data on study findings.

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 ensure that a registry includes data elements necessary for the intended use of the registry. Sponsors should confer with the appropriate FDA review division about the adequacy of registry data to support a regulatory decision, including the completeness and timing of data collection. The following are non-exhaustive examples of potential data that can be included in a registry:

- Demographic and clinical information:
  - Patient demographic factors, including date of birth, sex, sexual orientation, gender, race, and ethnicity, as well as measurements such as height and weight
  - Substance use, such as alcohol, recreational drugs, and tobacco
  - Primary diagnosis of interest, including date of diagnosis, diagnostic 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

### ***Contains Nonbinding Recommendations***

- 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
- Information on treatment of the disease of interest and comorbid conditions:
  - Drug chemical name and product name, formulation and dosage, start and end dates, and reason for discontinuation (as applicable) for each treatment
  - Type and date of procedures and periprocedural complications (as applicable)
- Other health-related information:
  - Specific clinical events of interest (e.g., heart attack, stroke, hospitalization, death) and date of occurrence
  - Other clinical outcomes (e.g., disease progression or relapse, disability, functional status, quality of life) and date of occurrence or assessment<sup>12</sup>
  - Pregnancy-related outcomes<sup>13</sup>

#### **C. Reliability of Registry Data**

When considering using data from an existing registry or establishing a registry de novo, sponsors should ensure that there are processes and procedures to govern registry operation, education and training of registry staff, resource planning, and practices that help ensure the quality of the registry data. Sponsors submitting data to FDA from registries held by another organization should work closely with that organization to ensure that processes and procedures are in place to support the reliability of the registry data. In addition, sponsors should have access to metadata associated with registry data intended to support regulatory decision-making.<sup>14</sup>

---

<sup>12</sup> See footnote 11.

<sup>13</sup> 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.

<sup>14</sup> For additional discussion of metadata, see the draft guidance for industry *Electronic Systems, Electronic Records, and Electronic Signatures in Clinical Investigations: Questions and Answers* (March 2023). When final, this guidance will represent FDA's current thinking on this topic.

## *Contains Nonbinding Recommendations*

Governance attributes that help ensure that the registry can achieve its objectives include, but are not limited to:

- An established data dictionary, which is available for those who intend to use the registry data to perform analyses. The data dictionary should generally include:
  - Data elements and corresponding definitions
  - Ranges, units, allowable values, and format for the data elements as well as corresponding data standards and terminologies used
  - Information regarding the origin of the data for each data element
- Defined processes and procedures for the registry, such as:
  - Collection, organization, maintenance, and storage of data, including processes to help ensure that data within a registry can be verified by source data (as applicable) for that registry
  - Documentation of (1) queries to ensure registry data are logically, consistently, and completely entered, including dates, times, and results of queries, and (2) steps taken to resolve issues identified, such as invalid, missing, or inconsistent data, including any changes made to the registry data
  - Prespecified rules for entering, modifying, searching, and retrieving registry data, including for situations in which registry data are combined with data from another source
  - Validation of the electronic systems used to collect registry data<sup>15</sup>
- 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<sup>16</sup>

Sponsors also should ensure that a registry adheres to applicable jurisdictional human subject protections requirements, including protecting the privacy of patient health information, when designing a registry or considering use of data from an existing registry. FDA also recommends

---

<sup>15</sup> Validation of electronic systems means a process of establishing and documenting that the specified requirements of an electronic system can be consistently fulfilled from design until decommissioning of the system or transitioning to a new system. This topic is also discussed in the draft guidance for industry *Electronic Systems, Electronic Records, and Electronic Signatures in Clinical Investigations: Questions and Answers*.

<sup>16</sup> For additional discussion on the use of electronic records and electronic systems under part 11, see the draft guidance for industry *Electronic Systems, Electronic Records, and Electronic Signatures in Clinical Investigations: Questions and Answers*.

## *Contains Nonbinding Recommendations*

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 whether the data were collected in accordance with the prespecified protocol. FDA also considers whether the registry personnel and processes in place during data collection and analysis are adequate to minimize errors and maintain integrity of the study data. Sponsors should address whether the registry has privacy and security controls in place to ensure that the confidentiality and security of data are preserved.<sup>17</sup>

Sponsors should use common data elements to promote standardized, consistent, and universal data collection. Such an approach can facilitate comparing or linking registry data to data from other sources. Standardized terminology and the associated data standards used by the registry should facilitate conformance with FDA's requirements for submitting study data in applicable drug submissions.<sup>18</sup>

Additional policies and procedures should be in place that enable FDA to assess the quality of the data (e.g., to address issues such as errors in coding or interpretation of source documents, data entry, cleaning, transfer, and linkage).

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

- Implement and maintain version control by documenting the date, time, and originator<sup>19</sup> 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 that data transferred from another data format or system are not altered in the migration process

---

<sup>17</sup> 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).

<sup>18</sup> See FDA's Study Data Standards Resources web page, available at <https://www.fda.gov/forindustry/datastandards/studydatastandards/default.htm>. For additional discussion of considerations for the use of data standards currently supported by FDA in applicable drug submissions containing study data derived from RWD sources, see the draft guidance for industry *Data Standards for Drug and Biological Product Submissions Containing Real-World Data* (October 2021). When final, this guidance will represent FDA's current thinking on this topic.

<sup>19</sup> Source data originators include persons, systems, devices, and instruments. For additional information, see the guidance for industry *Electronic Source Data in Clinical Investigations*.

### *Contains Nonbinding Recommendations*

- 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 collected over time (e.g., to align data elements with evolving diagnostic criteria or treatment guidelines)
- Explain the auditing rules and methods used and the mitigation strategies used to reduce errors
- Describe the types of errors that were identified based on audit findings and any corrective actions taken.

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 a digital health technology. This linkage may occur by populating the registry with data from another data system or by combining external data with registry data at the time of the analysis.

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 account for missing data, correct for redundant data, resolve any inconsistencies in the data, and 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 and accuracy 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 (e.g., data element mapping) to ensure appropriate 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) the data are accurately, consistently, and completely transmitted. Predefined rules to check for logical

### *Contains Nonbinding Recommendations*

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 from an external data source to the registry 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.<sup>20</sup> 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 linkage of records between the two (or more) databases can be performed accurately
- The linkage may compromise the representativeness and relevance of the patient population, if only a subset of the overall registry population is represented in the linked dataset
- The variables of interest in the registry and additional data sources have consistent definitions and reliable ascertainment approaches
- The data linked from external sources have been captured with sufficient accuracy, consistency, and completeness to meet study objectives
- The source data are available for any data sources used to supplement registry data
- The procedures and access controls in place protect the privacy of patient data
- Informed consent(s) provided by patients contributing data to each data source would allow for the linkage

After a sponsor decides to use an additional data source or sources to supplement the registry, the sponsor should develop and describe the approach used to link such data to the registry. FDA does not endorse a particular approach to linkage. The sponsor should also describe how data integrity will be evaluated after the linkage, including plans to assess any inaccuracies introduced by the linkage (e.g., overcounts of a particular data measure, duplication of patients or variables),

---

<sup>20</sup> For additional discussion on the use of electronic records and electronic systems under part 11, see the draft guidance for industry *Electronic Systems, Electronic Records, and Electronic Signatures in Clinical Investigations: Questions and Answers*.

## *Contains Nonbinding Recommendations*

steps taken to address issues identified, and approaches taken to address issues that cannot be fully rectified. When sponsors intend to submit registry data to FDA that include data from sources owned by another organization, sponsors should work closely with the other organization to ensure that appropriate methods for data entry, coding, cleaning, and transformation are in place for each linked data source.

### **E. Considerations for Regulatory Review**

Sponsors interested in using a specific registry as a data source for a study intended to support a regulatory decision by FDA should meet with the relevant FDA review division before conducting the study.<sup>21</sup> 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; and (5) the planned methods to validate the diagnosis of the disease being studied.

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.<sup>22</sup> All essential elements of the design, analysis, and conduct of a study using registry data should be predefined, and for each study element (e.g., eligibility criteria, exposures, outcomes, covariates), the protocol should describe how that element will be ascertained from the selected data source or sources.

Sponsors seeking to use registry data to support a drug's effectiveness and/or safety in a marketing application should ensure that patient-level data are provided to FDA in accordance with applicable legal and regulatory requirements.<sup>23, 24</sup> If the registry data are owned and controlled by third parties, sponsors should ensure that relevant patient-level data can be provided to FDA and that metadata and source records necessary to verify the RWD are made available for inspection, as applicable.<sup>25</sup>

---

<sup>21</sup> For example, sponsors can request a Type C meeting for non-interventional studies. See the draft guidance for industry regarding formal meetings with FDA, *Formal Meetings Between the FDA and Sponsors or Applicants of PDUFA Products* (September 2023). When final, this guidance will represent FDA's current thinking on this topic.

<sup>22</sup> For additional discussion, see the guidance for industry *Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products* (August 2023).

<sup>23</sup> See, e.g., 21 U.S.C. 355, 42 U.S.C. 262, and 21 CFR 314.50 and 601.2.

<sup>24</sup> *Ibid.*

<sup>25</sup> See 21 CFR 312.58.