

February 28, 2022

Dockets Management Staff (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 1061
Rockville, MD 20852

Re: Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products; Guidance for Industry; Availability (Docket No. FDA-2021-D-1146)

To the Food and Drug Administration:

The RWE Alliance appreciates the opportunity to comment on the draft guidance entitled “Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products” (referred to here as the “Draft Guidance”).¹ We are a coalition of real-world data (“RWD”) and analytics organizations with a common interest in harnessing the power of real-world evidence (“RWE”) to inform regulatory decision-making to improve patients’ lives. Our members have deep knowledge and experience working with healthcare data across disease areas and patient populations, and we aim to bring these collective insights to bear in support of RWE policies.²

The RWE Alliance envisions a future in which data generated in everyday clinical practice and everyday life through electronic health records (“EHRs”), administrative claims and billing data, product and disease registries, personal devices, wearables, and health applications will be used to generate evidence that complements clinical trial data to inform regulatory decisions. To achieve this goal, the RWE Alliance advocates for policies that will (1) advance FDA’s RWE Framework, (2) encourage the use of RWE to better understand treatment effects in underrepresented populations, (3) enhance opportunities for RWE organizations to consult with FDA, and (4) increase communication on the generation and use of RWE.³

We commend FDA for issuing the Draft Guidance as part of the Agency’s RWE Program, consistent with its mandate under the 21st Century Cures Act. In particular, we appreciate and agree with FDA’s recognition that registry data can be useful in regulatory decision-making. This letter provides our comments on specific sections of the Draft Guidance. For ease of reference, the headings for each subsection

¹ 86 Fed. Reg. 67956 (Nov. 30, 2021).

² For information about our members, please see our website, <https://rwealliance.org/who-we-are/>.

³ Additional information about what we believe is available on our website, <https://rwealliance.org/what-we-believe/>.

correspond to the titles used in the Draft Guidance: Section I provides recommendations to improve stakeholder understanding of registry data; Sections II and III highlight suggestions for evaluating the relevance and reliability of registry data, respectively; Section IV discusses considerations when registry data are linked to another RWD source; and Section V covers our comments on the Draft Guidance's glossary.⁴

I. Using Registry Data to Support Regulatory Decisions

We appreciate FDA's acknowledgement that each registry has unique attributes that will inform if it is suitable for use in regulatory decision-making.⁵ We agree with this overarching perspective on determining the suitability of registry data and encourage the Agency to make this point consistently across the RWE guidances.⁶

We also recommend enhancing the Draft Guidance's discussion of registries and registry data in the following ways to improve stakeholder understanding of this data source and its potential use in regulatory decision-making.

First, we recommend that FDA distinguish in the final guidance between the definitions of a "registry" and a "registry-based study" and add these terms to the glossary. For the definition of "registry," we recommend modifying the Draft Guidance's definition to clarify that a registry is an infrastructure for systematic data collection involving patients with a particular characteristic in common (e.g., a specific disease or treatment). The revised definition could describe a registry as "an organized system that uses observational study methods to collect uniform data (clinical and other) to evaluate specified outcomes for a population defined by a particular disease, condition, or exposure, and that serves one or more stated scientific, clinical, or policy purposes."⁷ In contrast, a registry-based study is conducted within that infrastructure using registry data. Registry-based studies may involve primary data collection, such as if the registry is a newly created infrastructure that is fit-for-purpose for a given research question or adds

⁴ Certain topics addressed in the Draft Guidance also are discussed in FDA's draft guidance entitled "Considerations for the Use of Real-World Data and Real-World Evidence To Support Regulatory Decision-Making for Drug and Biological Products." We plan to comment on overarching regulatory issues in the docket for the Considerations guidance given the guidance's more general applicability, but we strongly encourage FDA to consider feedback on these topics across the guidances.

⁵ Draft Guidance at Lines 129–132.

⁶ For example, we encourage FDA to clarify the draft guidance entitled "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products" by indicating that the Agency also views the suitability of RWD derived from EHRs and medical claims data as dependent on various factors, including the context in which the RWD would be used for regulatory decision-making.

⁷ See Gliklich RE, Leavy MB, Dreyer NA, AHRQ Publication No. 19(20)-EHC020, Registries for Evaluating Patient Outcomes: A User's Guide (4th ed. 2020), <https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-evaluating-patient-outcomes-4th-edition.pdf>.

to the scope of data collection within an existing registry; or a registry-based study may solely involve secondary data collection, such that all data are sourced from an existing registry.⁸

Second, we recommend refining the Draft Guidance’s definition of registries to indicate they collect “structured and *well-defined* data elements.”⁹ We recommend replacing “structured and well-defined” with “structured and *clearly defined*,” as “well-defined” has another meaning in the context of epidemiologic methods. We suggest that FDA also consider incorporating into this discussion the concepts of sufficient metadata quality, consistency, and completeness and how they facilitate the use of registry data in regulatory decision-making. The Agency could consider cross-referencing its recommendations on similar concepts in the draft guidance entitled “Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products” (“EHR/Claims Draft Guidance”). To the extent the recommendations in the EHR/Claims Draft Guidance should differ for registry data and registry-based studies, we recommend FDA clarify these distinctions in the final guidance.

Third, the Draft Guidance currently describes data heterogeneity as a “potential limitation[] of registries.”¹⁰ We recognize the importance of accounting for how heterogeneity may influence observed treatment effects. We encourage the Agency to provide recommendations on possible methods for addressing these issues in the final guidance, such as through data curation, ontological harmonization, and subgroup analyses. Furthermore, we recommend clarifying that there are additional considerations related to data heterogeneity that often arise when using multiple registries in a single study, such that pooling data or reconciling registry-specific results become necessary. More generally, however, we encourage FDA to expand this discussion of data heterogeneity in the final guidance to reflect that data heterogeneity—outside the context of pooling heterogeneous patient data that are not representative of the relevant population—can also present advantages. As discussed in our comments on the EHR/Claims Draft Guidance, heterogeneity in population characteristics can provide additional insight into diverse and underrepresented groups, which in turn can help advance the important public health objective of addressing disparities in healthcare access, treatment, and outcomes.¹¹

⁸ See European Medicines Agency, *Guideline on Registry-Based Studies* (Oct. 22, 2021), https://www.ema.europa.eu/en/documents/scientific-guideline/guideline-registry-based-studies_en-0.pdf. With respect to the latter example, greater clarity on FDA’s recommendations for research using existing registry data, such as when there are agreements between registry owners and third parties, would be beneficial.

⁹ Draft Guidance at Line 139.

¹⁰ *Id.* at Lines 152–153.

¹¹ See RWE Alliance, Comments on Docket No. FDA-2020-D-2307 9 (Jan. 24, 2022).

Fourth, the Draft Guidance states that “registries are better suited as a data source for regulatory purposes when sponsors aim to capture objective endpoints.”¹² We suggest clarifying that this discussion applies to endpoints such as: (1) conventional endpoints such as survival-based endpoints (e.g., overall survival and progression-free survival); and (2) validated surrogate endpoints that can be objectively assessed, such as plasma testosterone levels in advanced prostate cancer.¹³ We also recommend acknowledging in the final guidance that the general understanding of objective endpoints will likely evolve over time as new technologies, such as wearables, are increasingly used in patient care. The longitudinal collection of patient-reported outcomes and clinician-reported outcomes using validated instruments can be informative, particularly as measures of treatment response or other factors that change over time.

II. Relevance of Registry Data

We appreciate FDA’s recommendations on assessing the relevance of registry data.¹⁴ This evaluation should cover not just the data elements captured by the registry (as recommended in the Draft Guidance) but also other factors that may be applicable, such as the timing of when data elements are collected, missingness, and the version of coding used.¹⁵ We recommend clarifying in the final guidance that the assessment of a registry’s relevance for a particular research question should include consideration of (1) not only inclusion/exclusion criteria but also other factors that may influence a patient’s participation (e.g., disease severity or location of care) because all of these factors may affect the registry population’s representativeness of the target population; and (2) the extent and pattern of loss to follow-up (e.g., because patients no longer wish to participate in the registry, patients have transferred their care to a physician who is not a registry investigator, or they are lost for other reasons) because this factor is important for understanding the potential for missing outcome information.

In addition, although we understand that the Draft Guidance’s examples of potential data to include in a registry are non-exhaustive,¹⁶ we encourage FDA to consider adding biomarkers as a specific example of clinical information that may be included. Biomarkers—such as minimal residual disease status or the presence/absence of a genetic mutation or tumor type—serve critical roles in diagnosis, prognosis, treatment

¹² Draft Guidance at Lines 156–157.

¹³ See FDA, Table of Surrogate Endpoints That Were the Basis of Drug Approval or Licensure (Sept. 16, 2021), <https://www.fda.gov/drugs/development-resources/table-surrogate-endpoints-were-basis-drug-approval-or-licensure>.

¹⁴ We believe that a registry’s adequacy for evaluating scientific objectives is a consideration not only for using an existing registry but also for establishing a new registry. With this in mind, we suggest revising this section of the Draft Guidance to say, “When considering whether to use an existing registry *or to establish a new registry* for regulatory purposes” Draft Guidance at Line 175 (citation omitted) (emphasis added).

¹⁵ Similar to our comments in section I of these comments, FDA could consider referencing the aspects of data quality discussed in the EHR/Claims Draft Guidance.

¹⁶ See Draft Guidance at Lines 207–258.

selection, understanding the natural history of disease, and measuring disease activity such as response and progression. They can be accurate, valid, reproducible, and objective indicators of a patient's medical state. Moreover, the inclusion of biomarker data in registries is one factor distinguishing registries from other data sources that do not include biomarkers, such as medical claims. The inclusion of biomarker data in this list of examples would be consistent with FDA's explanation, earlier in the Draft Guidance, that registries can be used to identify biomarkers associated with important clinical outcomes that are relevant for study planning purposes.¹⁷

We also recommend providing the following clarifications in this section of the Draft Guidance:

- The Draft Guidance includes a recommendation with respect to using registry data as part of the enrollment process in an interventional study.¹⁸ We recommend updating this passage to refer not only to using a registry to identify eligible participants to enroll in an interventional study, but also to embedding the interventional study in a registry, where the existing registry infrastructure would be used to enroll patients and manage data collection throughout the study. Expanding this discussion accordingly would better reflect ways in which sponsors may use a registry for enrollment purposes.
- The Draft Guidance states, "The registry should retain information documenting any data elements that are no longer being collected or new elements that begin to be collected."¹⁹ We understand this statement to refer to data consistency and to recommend that sponsors and/or RWD/E organizations confirm that data elements are captured uniformly for the defined time period and patient population. We recommend the Agency introduce the term "data consistency" in this passage (rather than later on in the Draft Guidance) and update the sentence accordingly to help clarify this recommendation.
- We recommend refining the language regarding a "plan to reduce loss to follow-up of registry participants."²⁰ We believe this statement would provide more actionable guidance to stakeholders if it recommended developing "a plan to reduce loss to follow-up of registry participants and policies that define operationally how a registry participant is identified as lost to follow-up." We also recommend acknowledging in the final guidance (1) that the development of such policies and procedures may not always be possible, such as when a sponsor uses an existing registry from a third party and does not have influence on new data collection or follow-up; but (2) that the data derived from existing registry data can still be valuable and used in regulatory decision-making as long as the

¹⁷ See *id.* at Lines 105–107.

¹⁸ *Id.* at Lines 183–184.

¹⁹ *Id.* at Lines 184–186.

²⁰ *Id.* at Lines 186–187.

dataset is fit-for-purpose and the level of completeness of key variables is adequate to perform the analysis of interest (i.e., the loss to follow-up does not compromise the study's validity).

- The Draft Guidance states that “patients who remain enrolled in the registry may differ from those who do not remain (e.g., having experienced different adverse events).”²¹ We suggest replacing the phrase “do not remain” with “are no longer followed,” which is a term or classification more commonly used with respect to registries and would clarify which patients FDA is referring to here.
- We suggest revising “ultrasound reports that assess gestational age”²² to say instead, “gestational age estimates obtained from ultrasound reports, if available.” As written in the Draft Guidance, the recommendation could be misunderstood as suggesting that ultrasound reports themselves are always to be included as part of the registry data, which may not necessarily be the case.

III. Reliability of Registry Data

We strongly agree with FDA about the importance of ensuring the quality and reliability of registry data. We also agree that the use of common data elements can help promote standardized, consistent, and universal data collection in a registry.²³ We encourage FDA to update this passage to reflect the practical consideration, however, that the use of common data elements may not always be possible—e.g., if the same variables are measured or categorized differently across registries depending on the study population and/or the available data source. These lines could instead read, “When feasible, the use of common data elements can promote standardized, consistent, and universal data collection.” Clarifying this recommendation in this way would underscore the potential benefits of common data elements while also conveying that registry data can still be appropriate for regulatory purposes even when common data elements are not used.

Additionally, we recommend that FDA update the discussion of standardized terminology and associated data standards to reflect the incongruities between the Agency's Data Standards Catalog, on the one hand, and the data organization and formats in RWD sources. We appreciate that the Agency recognized these challenges in the draft guidance entitled “Data Standards for Drug and Biological Product Submissions Containing Real-World Data” (“Data Standards Draft Guidance”). Granularity in the source data may be lost when mapping the data to a currently supported format, and overemphasis on alignment with these standards could result in the exclusion of RWD sources that are otherwise deemed fit-for-purpose for a particular research question. As discussed in our comments on the Data Standards Draft Guidance, we encourage FDA to consider alternate approaches for reviewing RWD

²¹ *Id.* at Lines 195–196.

²² *Id.* at Lines 248–249.

²³ See *id.* at Lines 315–316.

(e.g., using an analytics platform or submitting RWD together with R or SAS scripts for Agency reviewers to run) that would retain the valuable insights found in source data without introducing the potential issues that can occur when converting an analytic dataset to a supported data standard.²⁴ We also agree with FDA’s ongoing efforts to update the Data Standards Catalog to include additional standards that are more compatible with RWD sources.²⁵ For example, in the context of registry data, we recommend adding support for standards that align with the data elements and class standards already specified in the Office of the National Coordinator for Health Information Technology’s (“ONC’s”) U.S. Core Data for Interoperability (“USCDI”).²⁶ We would welcome the opportunity to work with the Agency and other stakeholders to accelerate the process of identifying additional standards.

We also suggest clarifying this section of the Draft Guidance to address the following points:

- We recommend that FDA clarify the Draft Guidance’s reference to “rules for the validation of queries and edit checks of registry data.”²⁷ This language reflects clinical research terminology, such as “queries,” that may not be readily understood by stakeholders from other disciplines. We suggest clarifying that the use of “queries” in this context refers to questions sent by the registry staff to the investigator to clarify or confirm a specific data point. We also suggest clarifying that “edit checks” are rules that ensure the data are logically, consistently, and completely entered at the point of data entry and during data quality review.
- The Draft Guidance recommends defining processes and procedures for the registry, such as “plans for how *patients*, researchers, and clinicians will access and interact with the registry data and the registry’s data collection systems.”²⁸ To facilitate stakeholder understanding of this recommendation, we suggest that FDA clarify how it sees patients interacting with registry data, ideally using examples of such interactions.
- The Draft Guidance also recommends implementing policies and procedures “that enable FDA and persons interested in using the registry’s data to assess the quality of the data.”²⁹ We recommend clarifying that this objective can be satisfied by establishing means of assessing data quality (e.g., through statistical

²⁴ See RWE Alliance, Comments on Docket No. FDA-2021-D-0548 3 (Feb. 4, 2022) (“Comments on Data Standards Draft Guidance”).

²⁵ See *id.* at 3–4.

²⁶ To the extent the USCDI standards do not provide sufficient granularity for FDA’s purposes, we encourage the Agency to work with ONC to expand the USCDI standards as needed.

²⁷ Draft Guidance at Lines 268–269 (emphasis removed).

²⁸ *Id.* at Lines 278–279 (emphasis added).

²⁹ *Id.* at Lines 322–326.

characterization) without providing access to raw data when authorization for such access may not be feasible or such access is otherwise inappropriate.³⁰

- We ask that FDA clarify the last data management strategy included in the Draft Guidance, which recommends “[d]escrib[ing] the types of errors that were identified based on audit findings and how the data were corrected.”³¹ As written, this recommendation appears to focus on correcting individual data elements. We suggest revising the recommendation to say “[d]escrib[ing] the types of errors that were identified based on audit findings *and what corrective actions, if any, were taken.*”
- In addition, we ask FDA to provide examples of particular risk-based database quality assurance practices it considers appropriate for registries, such as those described by the Agency for Healthcare Research and Quality.³² Historically, many registries developed for health care quality improvement or use cases other than the development of new drugs or medical products have not invested in data governance boards and audit systems that FDA may wish to see going forward.

IV. Considerations When Linking a Registry to Another Registry or Another Data System

The Draft Guidance recommends that sponsors consider whether linked data sources are “interoperable.”³³ We encourage FDA to remove this reference to interoperability and recommend instead that sponsors (1) ensure data transfer and linking methods are appropriately tested, reliable, and accurate; (2) explain these methods in any regulatory submissions making use of these data; and (3) rely on established standards, such as those in the USCDI discussed above, that support linking whenever possible and appropriate (similar to using HL7 FHIR for EHR data). This approach would provide sponsors and other stakeholders with a practical recommendation to guide the process of linking systems. As noted in our previous comments, a well-designed and well-executed methodology should be the most important consideration for determining an effective approach to data linkage.³⁴ Emphasizing the specific technical means by which the systems are linked together could unduly limit which data sources can be

³⁰ It is unclear who FDA is referring to with the language “persons interested in using the registry’s data”—e.g., if it refers to third parties who are not involved with the registry (e.g., sponsors) or to someone else. We believe the Agency may be referring here to sponsors and suggest clarifying this point in the final guidance.

³¹ Draft Guidance at Lines 352–353.

³² See Gliklich RE, Leavy MB, Dreyer NA, AHRQ Publication No. 19(20)-EHC020, Registries for Evaluating Patient Outcomes: A User’s Guide (4th ed. 2020), Chapter 11. Obtaining Data and Quality Assurance (pages 290-314), <https://effectivehealthcare.ahrq.gov/sites/default/files/pdf/registries-evaluating-patient-outcomes-4th-edition.pdf>.

³³ See Draft Guidance at Lines 383–390.

³⁴ See RWE Alliance, Comments on Docket No. FDA-2020-D-2307 8 (Jan. 24, 2022).

linked and constrain potential research opportunities. Moreover, the two-way transfer of information often is not feasible in the registry context, so the term “interoperability” as used in the Draft Guidance is not particularly suitable as a key criterion.

In addition, we recommend that FDA clarify its recommendation for the sponsor to document the process that sponsors use to validate “the transfer of data.”³⁵ It is unclear which “transfer of data” FDA refers to in this passage: whether it is the transfer of data occurring within the RWD/E organization or for transfers of data from the RWD/E organization to FDA. Nor does the Draft Guidance specify which entity (e.g., the sponsor or the RWD/E organization) is responsible for retaining this documentation and for how long. We recommend that FDA provide clarity on these points.

We also suggest that FDA consider highlighting potential bias as an additional consideration for linking data sources. For example, selection bias could be a consideration if only a certain subset of patients is linked from a data source.³⁶ Consistent with other Agency recommendations,³⁷ we suggest that FDA add that sponsors evaluate whether and how bias, if present, might affect the study results.

V. Glossary

We appreciate the inclusion of a glossary in the Draft Guidance, as in the other RWE draft guidances published by FDA. We offer the following suggestions for the Agency’s consideration:

- The current definition of “data curation” refers to the “[a]pplication of standards (e.g., Clinical Data Interchange 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.”³⁸ As in our comments on the Data Standards Draft Guidance, we recommend updating this definition to remove the reference to CDISC, which is generally used for structuring data collected in clinical trials.³⁹
- We suggest expanding the Draft Guidance’s current definition of “reliability” in the main text to include “reproducibility,”⁴⁰ as well as adding “relevance” and “reliability” to the glossary because they are key terms in the Draft Guidance as well as for RWD/E more broadly. FDA could further enhance these definitions by clarifying how the Agency sees key concepts like relevance, reliability, and data

³⁵ Draft Guidance at Lines 392–393.

³⁶ The Draft Guidance appears to contemplate this potential for bias in section III.B of the Draft Guidance. See *id.* at Lines 189–196.

³⁷ See EHR/Claims Draft Guidance at Lines 488–490.

³⁸ Draft Guidance at Lines 466–469.

³⁹ See Comments on Data Standards Draft Guidance at 8.

⁴⁰ See Draft Guidance at Lines 170–171.

integrity as interrelating. For example, the current definition of “data integrity” refers to the “completeness, consistency, and accuracy of data”⁴¹ while the definition of “reliability” refers to “data accuracy, completeness, provenance, and traceability” (each of which are their own defined terms).⁴²

More generally, we recommend that FDA consider publishing a glossary that consolidates key terms across RWE guidances and harmonizes their definitions. Terms like “EHR” are defined in the EHR/Claims Draft Guidance, for example, but also appear in this Draft Guidance (as well as others). Establishing a glossary that centralizes FDA’s definitions would facilitate stakeholder understanding of the current suite of RWE guidances, and the glossary could be expanded in the future when the Agency publishes additional guidances on using RWE in regulatory decision-making.

VI. Conclusion

The RWE Alliance appreciates the Agency’s commitment to advancing the use of RWD and RWE in regulatory decision-making. Thank you for considering these comments, and please let us know if you have any questions. We would welcome the opportunity to discuss further.

Best regards,

The RWE Alliance

⁴¹ *Id.* at Line 473.

⁴² *Id.* at Lines 170–171.