REAL-WORLD EVIDENCE ALLIANCE

info@rwealliance.org
rwealliance.org

February 4, 2022

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

Re: Data Standards for Drug and Biological Product Submissions Containing Real-World Data; Draft Guidance for Industry; Availability (Docket No. FDA-2021-D-0548)

To the Food and Drug Administration:

The RWE Alliance appreciates the opportunity to comment on the draft guidance entitled "Data Standards for Drug and Biological Product Submissions Containing Real-World Data" (referred to here as the "Draft Guidance").[1] We are a coalition of real-world data ("RWD") and analytics organizations with a common interest in harnessing the power of real-world evidence ("RWE")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]

I. General Comments on the Draft Guidance

We thank FDA for issuing the Draft Guidance as part of the Agency's RWE Program, in partial fulfillment of the direction under the 21st Century Cures Act. We appreciate the Draft Guidance's recognition that data standards can apply differently to RWD as

---

[1] 86 Fed. Reg. 58672 (Oct. 22, 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/.

compared to clinical trial data, which are collected, standardized, validated, and submitted to the Agency using different methods and processes.

The following subsections provide our general comments on the Draft Guidance.

A. We encourage FDA to consider alternate approaches for reviewing RWD that would retain valuable insights found in the source data in its native format, though we also agree the Data Standards Catalog should be updated

We appreciate the Agency's statement recognizing "that a range of approaches may be used to apply currently supported data standards . . . to RWD sources such as EHR or claims data."[4] We agree there is not a one-size-fits-all approach for ensuring that RWD submitted to the Agency can be supported, and we encourage FDA to state it will apply flexibility in its discussions regarding specific drug submissions. The data formats in the Data Standards Catalog—particularly the Clinical Data Interchange Standards Consortium ("CDISC") Study Data Tabulation Model ("SDTM")—are designed to organize clinical trial data and do not consistently align with data organization and formats in RWD sources. As the Draft Guidance acknowledges, this lack of alignment can create challenges in converting RWD to a currently supported data format.[5] For example, granularity in the source data (and the insights that can be gleaned from this information) may be lost when mapping the data to a currently supported format,[6] or certain RWD sources may be excluded from use as a threshold matter due to a lack of alignment. These consequences may result in reduced internal validity and/or external validity (i.e., generalizability) for submitted studies. The linkage of data across multiple RWD sources can raise further considerations because each source may diverge from FDA's currently supported data standards in different ways.

We encourage FDA to take additional steps to address these considerations. As the Agency knows from its own independent analyses of raw data submitted in marketing applications, researchers have the most flexibility, the strongest understanding of study variables, and the most refined comprehension of underlying information when data are in their rawest form. Each new transformation constitutes an opportunity to introduce bias into an RWE study or to lose the rich detail captured in the original data. We provide the following recommendations with the intent of facilitating FDA's review of

---

[4] Draft Guidance at Lines 125–127.

[5] *Id.* at Lines 159–166. There may be differences in the data transformation process for prospective and retrospective studies. For an RWE study involving intentional data collection, there may be fewer challenges with capturing data fields that simplify mapping approaches. We suggest highlighting this distinction as another consideration in the final guidance.

[6] As an example, in one RWE study, conversion to CDISC SDTM not only resulted in additional investigator burden when mapping but also reduced the amount of information available in the data. *See* Garza, M., Del Fiol, G., Tenenbaum, J., Walden, A., & Zozus, M. N. (2016). Evaluating common data models for use with a longitudinal community registry. *Journal of biomedical informatics*, *64*, 333–341. https://doi.org/10.1016/j.jbi.2016.10.016.

regulatory submissions containing RWD and addressing the limitations presented by current data standards.

First, we encourage the Agency to consider new approaches for reviewing analytic datasets and results, particularly data from RWD sources, that would not require standardizing the analytic datasets to conform to a specific format in order to be submitted. For instance, it is possible to review RWD from a variety of sources on a single analytics platform, which would enable Agency reviewers to access source data and analyses in a transparent manner while maintaining a level of detail that might be lost in the process of standardizing the data to a currently supported format. Another option could be for an applicant to submit RWD together with R or SAS scripts that Agency reviewers can run to validate the analysis or even the underlying data.[7] Such alternate approaches could also help address other challenges, such as the possibility that a Define-XML file containing metadata for an RWE study would be very different from a file for a traditional clinical trial. The file for an RWE study may consist of pooled datasets from multiple institutional databases with dissimilar variables, which could render a Define-XML file very complex and thus difficult to interpret during the review process.

Second, expanding the Data Standards Catalog to include data formats that are more consistent with RWD sources, including when data from multiple sources are linked, would help address some of the limitations presented by currently supported data standards. The RWE Alliance supports FDA's proposal to do so, as described in the Draft Guidance,[8] and applauds the Agency's efforts to collaborate with stakeholders to evaluate the potential use of new data standards.[9] We would welcome the opportunity to work with the Agency and other stakeholders to accelerate the process for identifying these standards.[10]

In light of these ongoing efforts, we respectfully request that FDA expedite the updates to the Data Standards Catalog so any new standards are added before the Agency publishes a final version of the Draft Guidance. We believe this approach would be appropriate so the final guidance's recommendations can reflect the addition of new

---

[7] *See*, *e.g.*, R Consortium, *Successful R-Based Test Package Submitted to FDA* (Dec. 8, 2021), https://www.r-consortium.org/blog/2021/12/08/successful-r-based-test-package-submitted-to-fda.

[8] *See* Draft Guidance at Lines 108–109.

[9] *See* Reagan-Udall Foundation for the FDA, *Real-World Data Webinar Series: Data Standards for Drug and Biological Product Submissions Containing Real-World Data – Draft Guidance for Industry*, https://reaganudall.org/news-and-events/events/real-world-data-webinar-series-data-standards. We welcome additional public communications from FDA about these efforts to identify and develop additional standards and about the Agency's experience reviewing RWD.

[10] We also note the separate initiatives underway to promote data standardization, such as the Office of the National Coordinator for Health Information Technology's ("ONC's") framework for the sharing of health data between networks. *See* ONC, *The Trusted Exchange Framework (TEF): Principles for Trusted Exchange* (Jan. 2022), https://www.healthit.gov/sites/default/files/page/2022-01/Trusted_Exchange_Framework_0122.pdf.

data standards that are more compatible with RWD submissions, as envisioned by the Agency.

Additionally, the Data Standards Catalog should extend beyond SAS-based formats to include packages and modules developed using other programming languages (e.g., R, Python) or that are agnostic to programming language (e.g., JSON or XML formats). Incorporating RWD into clinical trials through adaptive designs, Bayesian analysis, or the use of more complicated machine learning models or imaging processing is more easily done—or in some cases can only be accomplished—outside of SAS. Identifying these additional standards for inclusion in the Data Standards Catalog would help facilitate appropriate and efficient data transformation, without compromising information about the underlying clinical context.

Similar issues arise for the Agency's other data standards resources that are not specific to RWD/E—e.g., FDA's Business Rules and Validator Rules to ensure data standard conformance in regulatory submissions. As an example, both sponsors and FDA use compliance checks to ensure that submissions of clinical trial results comply with FDA's Business Rules and CDISC standards. It is not clear, however, if FDA expects RWD submissions to undergo the same compliance checks. Such a process could be very burdensome for these analytic datasets due to the volume of data that may be involved and their incongruence with the compliance check's standards. If FDA expects compliance checks would be conducted for regulatory submissions containing RWD, we encourage the Agency to consider working with RWD/E organizations and sponsors to develop compliance checks that are tailored to these analytic datasets.

Third, we recognize that alternate approaches for reviewing RWD will take some time to develop fully, as will identifying new standards for the Data Standards Catalog. Therefore, in the nearer term, we would welcome:

- FDA's recommendations on the best process for an RWD/E organization to request a meeting with the Agency to discuss questions arising outside the context of an individual sponsor's submission, as mentioned in previous comments to the Agency.[11] These meetings could cover, for example, technical issues related to converting RWD to FDA-supported data standards across projects.

- Additional guidance from FDA regarding standards for study data derived from RWD, as proposed in the Draft Guidance.[12] In particular, we suggest providing guidance on applying supported data standards to linked data from multiple RWD

[11] *See* RWE Alliance, Comments on Docket No. FDA-2020-D-2307 (Jan. 24, 2022) ("Comments on EHR/Claims Draft Guidance"); RWE Alliance, Comments on Docket No. FDA-2021-N-0891 (Oct. 28, 2021).

[12] *See* Draft Guidance at Lines 108–109.

sources.[13] We would also welcome Agency engagement through public workshops addressing this topic on a periodic basis (e.g., biannually).

We agree with the importance of ensuring traceability for transformed data and hope to work with the Agency in identifying workable, transparent solutions.

B. We recommend clarifying that an RWD/E organization may have the direct insight necessary to address the Draft Guidance's recommendations regarding the curation and transformation of RWD

The Draft Guidance highlights various issues for sponsors to address regarding the conversion of RWD to a currently supported data standard. For example, the Draft Guidance recommends documenting the processes for managing RWD during data curation and data transformation,[14] the changes to data to conform to FDA-supported standards,[15] and the rationale for selecting particular CDISC data elements for an analytic dataset.[16]

Although the sponsor is responsible for the study itself, the RWD/E organization will typically have deeper visibility into some of the topics raised in the Draft Guidance because of the RWD/E organization's role in producing the final analytic dataset for a study. Certain steps in this process may implicate intellectual property and/or data privacy considerations for the RWD/E organization. For example, in some studies, an RWD/E organization is responsible for reviewing and transforming the source data, then producing a summary, de-identified analytic dataset for the sponsor. Under this scenario, a sponsor would not be directly involved in all of the steps before the final analytic dataset is delivered because it has arranged for the RWD/E organization to execute this process. In contrast, with a traditional prospective clinical trial, the sponsor often has full access to the underlying source documentation and can address specific topics directly with FDA. We suggest reflecting this distinction in the final guidance.[17]

C. We encourage FDA to help facilitate stakeholder understanding of how FDA's recommendations interrelate across RWD/E guidance documents

We greatly appreciate FDA's efforts to issue guidance addressing key topics for using RWD/E in regulatory decision-making and look forward to the Agency's publication of additional guidance documents in the future. Although the substantive focus of each RWD/E guidance is relatively distinct, many share common threads—e.g., because they

---

[13] Guidance on evaluating regulatory considerations for linked data would also be welcome.

[14] *Id.* at Lines 98–101.

[15] *Id.* at Lines 101–102.

[16] *Id.* at Line 147.

[17] We appreciate that FDA has provided recommendations about data access in more recent draft guidances, and we will be submitting separate comments to those dockets. *See*, *e.g.*, FDA, *Draft 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* 6 (Dec. 2021).

touch on the same phases of the data life cycle or address similar considerations. We encourage FDA to take additional steps to clarify how the various guidance documents and their substantive recommendations interrelate.

As an example, section III.B of the Draft Guidance ("Documentation of Processes for Managing Real-World Data") could reference section VI of the separate draft guidance entitled, "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products" (referred to herein as the "EHR/Claims Draft Guidance"), which discusses additional considerations related to ensuring data quality. We offer additional suggestions in the comments below: see sections II.A, regarding the documentation of processes for managing RWD; II.B, regarding questions about applying the Draft Guidance's recommendations in light of the EHR/Claims Draft Guidance's discussion of common data models ("CDMs"), and II.D, regarding the Draft Guidance's glossary.

Another option could be to establish a centralized and public RWE Program "dashboard" on FDA's website that includes the most recent guidance documents related to RWD/E, among other information. We recognize FDA has already begun taking steps to do this and encourage the Agency to expand the information collected on its existing webpage.[18] In addition to the information currently provided, a dashboard could communicate information about actual RWD/E use cases, pilot projects, and the Agency's other collaborative efforts related to RWD/E, as we have proposed in separate comments.[19] FDA could also use the dashboard to aggregate relevant guidance documents as they are published and share links to other resources that the Agency views as relevant.[20]

II. Comments on Specific Sections of the Draft Guidance

The following subsections provide our comments on specific sections of the Draft Guidance. For ease of reference, the headings for each subsection correspond to the titles used in the Draft Guidance.

In addition, we also offer the following additional comment on the Draft Guidance that is not specific to any of its current sections: We would appreciate guidance about the applicability of FDA's current data standards to raw RWD submitted to the Agency at its request (e.g., data in DICOM format for medical imaging data, rather than the independent reader assessments that may be captured in a structured data field).

---

[18] *See* FDA, Real-World Evidence, https://www.fda.gov/science-research/science-and-research-special-topics/real-world-evidence.

[19] *See* Comments on EHR/Claims Draft Guidance.

[20] The NCI Thesaurus is an online database that provides reference terminology for various systems, for example, and a link to this resource could be shared on the dashboard.

A. Documentation of Processes for Managing Real-World Data

We strongly agree with FDA about the importance of implementing processes to increase confidence in the RWD used in a study. We would appreciate further guidance about the Agency's recommendations for any assessments that should be conducted to evaluate the final analytic dataset's validity—e.g., to confirm that any changes to the data to conform to FDA-supported data standards have been addressed appropriately—and to evaluate the potential impact of mapping RWD variables to conform with CDISC terminology.[21] If the EHR/Claims Draft Guidance's recommendations regarding validation and documentation apply in this context, the Agency should consider clarifying this point in the final guidance. We also recommend providing an example, either in this section or in the appendix, of a scenario in which mapping RWD variables to CDISC terminology would negatively affect the utility of the analytic dataset submitted for FDA's review.

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

In order to replicate analyses used to support a regulatory submission, sponsors converting data in accordance with the Draft Guidance would need to map RWD to a supported data standard prior to any analysis (e.g., prior to the creation of the analytic dataset). Such a mapping exercise would implicate many of the same considerations for CDMs described in the EHR/Claims Draft Guidance.[22] We would appreciate further guidance about applying the EHR/Claims Draft Guidance's recommendations for CDMs, such as when customized study-specific data elements are used to address deficiencies.

These recommendations should also address CDMs that are not identified in the Data Standards Catalog, as well as scenarios in which an analytic dataset uses variables from multiple CDMs or when additional variables may need to be created outside of those CDMs (or outside of the CDISC terminology). CDISC does not always allow for appropriate and relevant RWD elements to be captured, as noted in our general comments above. In contrast, individual CDMs—such as i2b2, Sentinel, HL7-FHIR, PCORNet, and others—and domains across different CDMs are often more suitable for use in an individual RWE study. We recommend that FDA consider (1) identifying these, or additional, CDMs as an option for standardizing data included in regulatory submissions; and (2) allowing data standardization through the use of the NIH BRIDG data model, through which some of the previously mentioned CDMs are mapped to specific CDISC domains.

---

[21] *See* Draft Guidance at Lines 102–103 (recommending the documentation of potential impacts from data changes to conform to FDA-supported data standards).

[22] *See* EHR/Claims Draft Guidance at Lines 329–343 (noting, for example, that data in CDM-driven networks "rarely contain all of the source information present at the individual healthcare sites").

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

The Draft Guidance recommends that sponsors "include a data dictionary that documents the definition of every data element used and all relevant information about the element . . . ."[23] We suggest enhancing this recommendation with details clarifying what a data dictionary should include. For example, we would appreciate a more detailed discussion of what FDA views as "relevant" information about a data element. We also recommend delineating when information about an RWD source's characteristics belongs in the data dictionary, as part of the comments section within CDISC and its dependencies, or as part of the Study Data Reviewer's Guide.

D. Glossary

Ensuring alignment across the defined terms in each RWD/E guidance will be critical to facilitating stakeholder understanding of how recommendations in the various documents interrelate. We recommend providing the same definitions for key terms across guidances to the extent possible, as recommendations in multiple guidances could apply to the same topic. As one example, the definitions of "data curation" are similar but not the same in the Draft Guidance and the EHR/Claims Data Draft Guidance. Whereas the Draft Guidance defines "data curation" as including the application of CDISC to source data,[24] the EHR/Claims Draft Guidance does not mention CDISC in its definition of the same term.[25] We recommend updating the definition of "data curation" in the Draft Guidance to align with the EHR/Claims Draft Guidance by removing the reference to CDISC, as this data standard is generally used for structuring data collected in clinical trials and FDA may support other standards in the future that are more compatible with RWD sources. We also recommend (1) adding the Draft Guidance's definition of "mapping" to the glossary in the EHR/Claims Draft Guidance, and (2) adding the EHR/Claims Draft Guidance's definitions of "data element" and "Common Data Model" to the Draft Guidance. These terms are used in each of the guidances, and expanding the respective glossaries to include them would help stakeholders better understand how the documents and the Agency's recommendations relate to one another.

We also suggest that FDA consider adding the following additional terms to the Draft Guidance's glossary.

- ***Ontology.*** The Draft Guidance includes terms such as "controlled terminology, "data standards," "mapping," and "terminologies." Adding "ontology" as it relates to information technology would be useful because some of the topics covered in the Draft Guidance relate to the use of ontologies, such as RxNorm (which is also a defined term in the glossary). We suggest briefly explaining the relevance

[23] Draft Guidance at Lines 151–152.

[24] *Id.* at Lines 185–188.

[25] EHR/Claims Draft Guidance at Lines 1300–1302.

of ontologies in the final guidance to connect these concepts. For the definition of "ontology," we propose adapting the following description: "Ontology . . . describes the concepts of medical terminologies and the relation between them [to] enabl[e] the sharing of medical knowledge."[26]

- ***SNOMED CT.*** We suggest adding a reference to SNOMED CT in the appendix to the Draft Guidance, at line 338 where RxNorm is mentioned. "SNOMED CT" could be defined in the glossary as "a comprehensive clinical terminology that provides clinical content and expressivity for clinical documentation and reporting."[27]

III. 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

[26] Aldosari B, Alanazi A, Househ M. Pitfalls of Ontology in Medicine. Stud Health Technol Inform. 2017;238:15-18. PMID: 28679876.

[27] NIH National Library of Medicine, *SNOMED CT FAQs*, https://www.nlm.nih.gov/healthit/snomedct/faq.html.