February 4, 2022

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

**Re: Docket No. FDA–2021-D-0548: Data Standards for Drug and Biological Product Submissions Containing Real-World Data**

Dear Sir/Madam:

The Biotechnology Innovation Organization (BIO) thanks the Food and Drug Administration (FDA or Agency) for the opportunity to submit comments regarding the draft guidance on **Data Standards for Drug and Biological Product Submissions Containing Real-World Data.**

BIO is the world's largest trade association representing biotechnology companies, academic institutions, state biotechnology centers and related organizations across the United States and in more than 30 other nations. BIO's members develop medical products and technologies to treat patients afflicted with serious diseases, to delay the onset of these diseases, or to prevent them in the first place.

BIO appreciates the opportunity to review and comment on the draft guidance. Below are a few overarching recommendations to improve the draft guidance:

1. **Considerations for Prospective and Retrospective Real-World Data**

Real-world data can be prospective or retrospective, and they have different strengths and limitations and the considerations for mapping each respective data source to CDISC standards may be different. For prospective real-world data (RWD) collection, it may be easier to have a good understanding of data generation, mapping and transformation; however, there are still challenges. For retrospective real-world data, in most circumstances, the Sponsor purchases data from data vendors. It can be challenging for the data vendors to change practice models based on a specific Sponsor's needs and standards. This issue is not taken into consideration in this draft guidance document. BIO recommends that the Agency consider highlighting the strengths and limitations of prospective and retrospective real-world data and providing language to delineate the challenges more clearly with each type of data source. BIO also recommends that the Agency consider highlighting the challenges between primary data (i.e., collected for research purposes) and secondary data (i.e., existing data such as claims or EHR) as conformance to FDA data standards will be much more challenging with secondary data.

2. **Considerations for Mapping and Transforming Real-World Data**

BIO appreciates that the Agency acknowledges the challenges in mapping and transforming RWD into data that meet FDA-supported data standards. BIO recommends that the Agency consider incorporating the following elements into the draft guidance:

- How to document the impact of data mapping and transformation on the source data.

- Detailed guidance on what is the acceptable level of discrepancies between FDA standards/terminologies and those for RWD.
- Detailed guidance on how to address data sources that may have used non-structured data (e.g., Natural Language Processing) to generate data elements.
- An acknowledgment of the issues/challenges raised when multiple RWD sources are used/linked and recommendations on how to address divergences among multiple data sources and FDA-supported data standards.
- Detailed guidance on data compliance. For example, in clinical studies Pinnacle 21 checks are executed before submission. It would be helpful for the Agency to clarify if RWD will be held to the same standard.
- Recommended best practices for handling RWD fields that have been auto-populated (Some EMR systems will automatically pull forward data from past visits into subsequent visits unless deliberately overwritten, and there may not be a flag for such auto-populated data). Recommendations on the kinds of data manipulation, if any, permitted under these circumstances would be helpful.

BIO also recommends that the Agency consider convening the relevant stakeholders that use RWD to discuss how to make these elements actionable and to provide clear guidance.

### 3. Considerations for Regulatory Flexibility for Data Standardization

While the FDA is applauded for seeking guidance for developing data standards to study data derived from RWD sources, the push to apply current FDA data standards to evolving and emerging data sets is somewhat challenging. Mainly, the data standards of existing data sets were created when the healthcare data ecosystem was not as diversified, heterogeneous, and technologically advanced. (e.g., https://jamanetwork.com/journals/jama/fullarticle/1883026). While these advancements in RWD types can pose challenges to data standardization, they also present opportunities for the FDA to re-evaluate data standards for existing data sources, and to collaborate with clinical researchers, data scientists, bioinformatics professionals, policy makers, and technologists to better organize data standards for RWD use in the 21st Century. The current guidance seeks to align modern data sets to data standards (those existing within the FDA) created for a different era in healthcare research. BIO recommends that the Agency consider providing guidance for a more far-reaching approach to data standards and create a new data standard approach tailored for RWD, given the evolving and emerging nature of RWD assets that can be at the FDA's disposal to facilitate data use for 21st Century research problems. One step in this direction may be to consider using RWD standards such as OMOP/CDM which are more flexible. BIO recommends that the Agency consider convening sponsors, data scientists, bioinformaticians, etc. to work together to address the aforementioned challenges. It may be worthwhile to consider workshops and/or pilot projects to have dialogue and consider other appropriate data standards for RWD.

BIO agrees that data standardization is a worthwhile goal for the use of RWD in the evaluation of product benefit/risk. Rigid requirements to transform all data into a format that was designed for clinical trials, however, may not be practical, scientifically appropriate, or feasible. For example, size may become an issue with mapping RWD to SDTM as many RWD studies have much larger cohorts than found in clinical trials and some of the mandatory/required SDTM fields are not available in RWD. BIO recommends that the Agency provide an appropriate degree of flexibility and guidance for when it is not feasible for the data to be standardized into the two currently recommended formats in the Catalog. For example, the use of RWD for the

study of rare diseases is very common as it is often not feasible or ethical to conduct RCTs in these populations. However, currently there are no standards for mapping of rare disease data to SDTM.

### 4. Harmonization of RWD/E Guidance

BIO recommends a greater degree of harmonization across the contemporary guidance documents on RWE. This harmonization would ensure that the Sponsor is incorporating all of the necessary information into the submission package to evaluate benefit risk. For example, the recent draft guidance "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products" outlines considerations with respect to the submission of data using common data models, yet this draft guidance does not contain any information on the use of common data models. Hence, it is not clear if the Sentinel common data model would be an acceptable format for submission of RWD.

As per above, the Agency should consider an opportunity to include details outlined in this draft guidance document in the data management plan described in "Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products**".** Coordinating and consolidating RWD submission documents would decrease burden on Agency review and incorporate less burdensome processes for the Sponsors.

Sincerely,

/s/
Camelia Thompson, Ph.D.
Senior Director, Science and Regulatory Affairs
Biotechnology Innovation Organization

**SPECIFIC COMMENTS**

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| **I. INTRODUCTION AND SCOPE** | | |
| **Lines 34-35** | "This guidance addresses considerations for the use of data standards currently supported by FDA in applicable drug submissions containing study data derived from RWD sources."<br><br>The scope of "drug submissions containing data derived from RWD sources" may range from (1) a few RWD variables collected in an interventional trial (e.g., hybrid trial), (2) a concurrent RWD study to serve as an external control arm to a single arm clinical trial, to (3) a study entirely based on RWD sources conducted either prospectively or retrospectively. Should data standard considerations be different in various scenarios? | BIO recommends the Agency provide further guidance on how these distinct scenarios of RWD in drug submissions will play a factor in selecting the appropriate data standard. In the first two scenarios, there may be a need to harmonize with data standards applied to the interventional portion of the study. Further, in scenario three, RWD collected prospectively vs retrospectively may impact the appropriate data standard used and new data standards more appropriate for RWD studies may be best to use in this scenario.<br><br>BIO recommends that the Agency clarify if clinical trial data considered for submission should be the guiding principle for RWD/RWE/Claims/Registry data as well. |
| **Lines 24-27** | The draft guidance states, "'….FDA has created a framework for a program to evaluate the potential use of real-world data (RWD) to generate RWE to help support the approval of new indication(s) for drugs3 already approved under section 505(c) of the FD&C Act (21 U.S.C. 355(c)) or to help support or satisfy post-approval study requirements (RWE Program).4'"<br><br>The introduction of the draft guidance states that it covers RWD/RWE to support the approval of new indication(s) for approved drugs or to satisfy post-approval study requirements. In Section II, the Agency referenced its June 2021 guidance *Providing Regulatory Submissions In Electronic Format — Standardized Study Data* (Study Data Guidance) | BIO recommends that FDA clarify in the final guidance whether the scope of this guidance applies to RWD/RWE used to support original NDAs. |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | which seems to imply that this guidance would able to original NDAs as well. Does this mean RW data/evidence cannot be used to support NDAs? | |
| **Lines 35-40** | The draft guidance states, "For the purposes of this guidance, FDA defines *RWD* as data relating to individual patient health status or the delivery of health care routinely collected from a variety of sources. Examples of RWD include data from ***electronic health records*** (EHRs); ***medical claims data***, data from product and disease ***registries***; patient-generated data (including data from in-home-use settings); and data gathered from other sources that can inform on health status, such as mobile devices."<br><br>This guidance discusses data standards requirements. The major challenge is standardization. | With all the variations of data collection, BIO recommends that the Agency clarify if there is a plan to standardize this so that downstream data/analysis can be standardized from a consistent format.<br><br>BIO also recommends that the Agency clarify whether eCOA or other data collected from mobile devices are in scope. |
| **II. REGULATORY BACKGROUND** | | |
| **Entire Section** | The use of RWE for FDA approval of drugs and biologics is not new. Evidence suggests that in oncology, FDA has used RWE for approvals:<br><br>Feinburg, et. al. (2020). https://www.sciencedirect.com/science/article/pii/S1098301520322026<br><br>However, in the review of these approvals, there is little to no insight provided into the details of data standards used to support these approvals. | BIO recommends that the Agency assess the RWD details of these approvals to begin providing enhanced guidance, based on real world research experience. |
| Lines 52-63 | The draft guidance states, "Under section 745A(a) of the FD&C Act, at least 24 months after the issuance | BIO recommends that the Agency clarify if RWD would adhere to the same standards as the Study Data Guidance |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | of a final guidance document in which FDA has specified the electronic format for submitting certain submission types to the Agency, such content must be submitted electronically and in the format specified by FDA.8 The guidance for industry, *Providing Regulatory Submissions In Electronic Format — Standardized Study Data* (Study Data Guidance), and the technical specifications referenced therein describe electronic submission requirements under section 745A(a) of the FD&C Act for clinical and nonclinical study data contained in new drug applications (NDAs), abbreviated new drug applications (ANDAs), certain biologics license applications (BLAs), and certain investigational new drug applications (INDs) submitted to the Center for Drug Evaluation and Research or the Center for Biologics Evaluation and Research.9 Given that these electronic submission requirements apply to study data submitted in the covered application types, they apply to RWD that is submitted as study data in such applications." | and thus already fall into the 2 year period or does the issuance of this final guidance start the 2 year period. |
| **Lines 68-70** | The draft guidance states, "…that use the standards specified in the Data Standards Catalog (Catalog)". | BIO recommends the Agency consider adding RWD standards to the Data Standards Catalog to account for nuances of RWD. |
| **III. APPLYING CURRENTLY SUPPORTED DATA STANDARDS TO STUDAY DATA DERIVED FROM REAL-WORLD DATA SOURCES** | | |
| ***A. Challenges in Real-World Data Standardization*** | | |
| **Entire Section** | The draft guidance identifies four challenges in real-world data standardization but not all of these are relevant to both retrospective and prospective data collection. It would be beneficial to note this in this section. BIO suggests that the Agency consider | BIO recommends the following edit:<br><br>**"...(5) RWD sources tend to have more sporadic data collection compared to a clinical trial, and can be longitudinal in nature often due to variation in data** |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | adding other fundamental challenges to this section but also noting that some of these obstacles can be overcome by designing appropriate prospective observational studies and using intentional data collection.. BIO also suggests that the Agency provide recommendations/suggestions on acceptable methods to reconcile the challenges.<br><br>The weight of evidence that the results from an RWD study provide will vary from application to application. In cases where the RWD study is serving as the "adequate and well controlled study" or is a key supportive study (such as a formal external comparator/control), datasets for the RWD study would appear to be required. Whereas, if a RWD study is being used in a very minor supportive way, such as to provide background information on treatment patterns, it would be reasonable not to include datasets.<br><br>The draft guidance discusses several challenges of standardizing data derived from RWD sources for inclusion in applicable drug submissions including multiple variations of RWD sources and inconsistent formats, and differences in source data, terminologies, and exchange formats. The agency states that it will provide guidance on RWD standards, but it implies that RWD will impact current data standards. | **collection time points from patient to patient, and (6) RWD data are not collected primarily for purpose of research, but rather for other reasons such as claims payment or a physician's own notes."**<br><br>BIO also recommends that the Agency provide recommendations/suggestions on acceptable methods to reconcile the challenges.<br><br>BIO recommends that the Agency clarify the scope of scenarios for RWD studies in which datasets are necessary and the required dataset standards are expected to be followed.<br><br>BIO recommends that the Agency consider additional guidance on RWD standardization first so that the RWD can be mapped appropriately. |
| **Line 86** | There may be different challenges encountered depending on the type of data source from which the RWD is derived, but it is not clear from the draft guidance whether there may be different expectations. This perhaps could be addressed | BIO recommends that the Agency provide additional guidance on the expectations of different data sources. |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | through discussions with the appropriate review division. | |
| **Line 93** | The draft guidance states, "…and de-identification methodologies used to protect patient data when shared" | BIO recommends the Agency provide a reference for acceptable de-identification methodologies, such as https://www.hhs.gov/hipaa/for-professionals/privacy/special-topics/de-identification/index.html |
| ***B. Documentation of Processes for Managing Real-World Data*** | | |
| **Entire Section** | The draft guidance states, "During data curation and data transformation, adequate processes should be in place to increase confidence in the resultant data. Documentation of these processes may include but are not limited to electronic documentation (i.e., metadata-driven audit trails, quality control procedures, etc.) of data additions, deletions, or alterations from the source data system to the final study analytic data set(s). Sponsors should also document in their applicable drug submission changes to data to conform to the current FDA-supported data standards, and the potential impacts of these changes."<br><br>Guidance within this document is limited regarding the documentation needed during data curation and data transformation steps. The Agency does not provide any guidance on which processes for managing RWD are mandatory/optional in documentation.<br><br>Sponsors typically partner with data providers to obtain access to RWD and based on contractual/legal reasons may not be able to obtain source data/information, which instead sits with the data provider. This may include information on data curation and transformation | BIO recommends the Agency reference the recent RWD guidance titled "Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products" within this section and more generally within this guidance.<br><br>BIO also recommends that the Agency provide more instruction on which processes for managing RWD are mandatory/optional in documentation.<br><br>The process of data curation is very technical; thus BIO recommends that the Agency collaborate with the Office of the National Coordinator (ONC) as it organizes technical standards for the use of technological tools (e.g., Application Programming Interfaces (APIs), Fast Healthcare Interoperability Resources (FHIR) and other tools that will be standards based to support the emerging form of RWD that are digitally created.<br><br>As this section discusses the need for such documentation, BIO recommends that the guidance acknowledge that Sponsors may be working with various partners for this information and that challenges/mitigations related to information access may exist but could be part of the Sponsor's discussions with the review division. |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | processes/documentation performed by the data provider. | Further, some data elements are often derived by the data provider (for example, overall survival, progression free survival, etc.) without the Sponsor having access to the source data in which the element was derived. It would be helpful for the guidance to clarify the expectations for such situations. As derived data, a logical consideration would be to incorporate this into analysis (i.e. ADaM) datasets. However, the guidance also specifies the need for SDTM datasets, which would involve the source data which the Sponsor may not have access to.<br><br>In the draft guidance, the Agency specifies that Sponsors should include information in several locations, including the Study Data Reviewer's Guide and the Define-XML/Domain files. A Data Dictionary is also requested to be included. Such a dictionary, in practice, is often specified in the Define-XML file. BIO recommends that the Agency clarify its expectations for the Data Dictionary, including whether the data dictionary must be an additional document or if it can be incorporated within the Define.xml. |
| **Lines 97-102** | The draft guidance states, "During ***data curation*** and ***data transformation***, adequate processes should be in place to increase confidence in the resultant data Documentation of these processes may include but are not limited to electronic documentation (i.e., metadata-driven audit trails, quality control procedures, etc.) of data additions, deletions, or alterations from the source data system to the final study analytic data set(s)." | Electronic documentation should include all of the alterations to source data. BIO recommends that the Agency clarify what additional documentation is needed for data curation and data transformation. |
| **Lines 101-103** | The draft guidance states, "Sponsors should also document in their applicable drug submission changes to data to conform to the current FDA- | BIO recommends that the Agency clarify where Sponsors should document changes to data to conform to the current FDA-supported data standard (i.e., should these be |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | supported data standards, and the potential impacts of these changes."<br><br>It is not clear where Sponsors should document changes to data to conform to the current FDA-supported data standard. | documented in the SDTM Reviewer's Guide and/or is there another forum for this?). |
| ***C. Considerations for Conforming Real-World Data to Currently Supported FDA Study Data Standards*** | | |
| **Line 109-112** | Currently, only SDTM and ADaM data formats are supported in the Catalog. | BIO recommends the Agency seek Industry feedback in updating the Catalog with additional standards for data using RWD sources. BIO also believes the Agency could describe the process for contacting the Agency if a Sponsor has a suggested addition to the Catalog and provide details on the amount of time it will take for a RWD-related data standard to be added. |
| **Entire Section** | It is acknowledged in the draft guidance that FDA is planning to issue further guidance and/or update the Catalog with standards for study data that are derived from RWD sources.<br><br>On one hand, the Agency is saying this document will provide guidance for RWD standards. On the other hand, it seems to be saying that RWD will also impact current standards. We hope that the standards would come first so the RWD data can be mapped appropriately.<br><br>When considering appropriate data standards for RWD, ADaM like principles could be applied and modified to appropriately capture RWD. | BIO recommends the Agency work closely with key stakeholder groups, including data standards consortiums, data providers, industry, and other health authorities, to help advance harmonized standards for RWD.<br><br>BIO suggests the Agency clarify if the comments from this guidance will help to address the Data Standards Catalog.<br><br>BIO suggests the Agency reference and discuss recent work on challenges and solutions for programmers dealing with RWD non-interventional studies. For example, PHUSE EU recently performed an assessment of challenges and various solutions for programmers when dealing with RWD non-interventional studies (https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Data+Standards+for+Non-interventional+Studies.pdf). This type of assessment and |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | | learnings from the PHUSE EU paper could be referenced and discussed in the guidance.<br><br>BIO recommends that the Agency first evaluate current data standards to assess how they apply to the emerging and evolving RWD ecosystem. Richesson, et. al. (2014): https://academic.oup.com/jamia/article/14/6/687/750453 recommend an assessment of existing data standards.<br><br>BIO recommends that the Agency provide additional guidance on data standardization needs based on "fit for purpose" of RWE study types and its regulatory needs with existing data standards supported by FDA. FDA is encouraged to assess how the current data standards conform with RWD used in studies such as pragmatic clinical trials, external control study, long-term follow-up studies, the use of artificial intelligence (AI) and machine learning (ML) to identify patients for new populations, new indications, and other types of cost effectiveness studies that will be used by regulatory agencies. Andre, et. al. (2019): https://onlinelibrary.wiley.com/doi/full/10.1002/pds.4932 |
| **Line 126** | The draft guidance currently seems to be limited to accepting CDISC SDTM data. | Similar to our comment on lines 109-112, BIO recommends the Agency consider other common data models such as Sentinel CDM or OHDSI OMOP which are more suitable to observational data and standardized across healthcare research in both academic and industry settings. [Reference: Maryam Garza, Guilherme Del Fiol, Jessica Tenenbaum, Anita Walden, Meredith Nahm Zozus, Evaluating common data models for use with a longitudinal community registry, |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | | Journal of Biomedical Informatics,Vol 64,2016,p333-341,ISSN 1532-0464,https://doi.org/10.1016/j.jbi.2016.10.016. (https://www.sciencedirect.com/science/article/pii/S1532046416301538)]<br><br>BIO also recommends that since the Sentinel CDM is used and supported by FDA, it would be helpful for the Agency to provide guidance based on its experience with Sentinel and fit-for-purpose use cases, e.g., https://www.sentinelinitiative.org/sites/default/files/documents/ICPE%20Presentation%20-%20Strategies%20for%20the%20Use%20of%20RWD%20to%20Conduct%20COVID-19.pdf |
| **Line 115** | There will be challenges with converting RWD into data consistent with FDA-supported data standards. | BIO recommends the Agency comment on and/or reference in the final guidance the recently published FHIR to CDISC Joint Mapping Implementation Guide by CDISC (01 September 2021) and their views for accepting this as an approach to remap data from EHR to CDISC SDTM variables. (https://www.cdisc.org/standards/real-world-data/fhir-cdisc-joint-mapping-implementation-guide-v1-0 ) |
| **Lines 115-117** | The draft guidance states, "When seeking to conform RWD to data standards supported by FDA, Sponsors should consider the relevant data transformations, conversions, or ***mappings*** that may be needed to produce study datasets in the required format in an applicable drug submission." | BIO recommends the Agency clarify if this means that technical rejection criteria is applicable to RWD studies. Also, if it is applicable, apart from TS, the agency should clarify if there is a need for other trial domains for RWD. |
| **Lines 119-121** | The draft guidance states, "Sponsors should discuss early, with the appropriate FDA review division, any planned submission of study data derived from RWD sources in an applicable drug submission and their approaches for transforming the data to the current FDA-supported data standards." | BIO recommends that the Agency provide additional guidance on the appropriate forum for Sponsors to discuss data standards with the Agency.<br><br>BIO also recommends that the Agency clarify when discussions for planned submissions of study data derived |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | It is not clear from the draft guidance when discussions with the appropriate FDA review division should occur. | from RWD sources should take place (e.g., as the protocol is in development or after approval). |
| **Lines 121-123** | The draft guidance states, "Sponsors should describe these approaches, including in the protocol, data management plan, and/or final study reports."<br><br>Often the Sponsor does not have access to the data at the time of protocol submission to be able to provide sufficient detail on the approaches. | BIO recommends that the Agency provide additional guidance on how to handle RWD used in analysis that is only available in the form of aggregate (i.e., the estimates of functional parameters) from registries and individual subject data is not available.<br><br>BIO recommends that the Agency clarify the level of detail requested in each of these documents. |
| **Line 125** | The draft guidance does not mention choices of approaches to apply CDISC to RWD. It would be of value to provide more information whether there are preferred approaches or approaches to avoid. | BIO recommends that the Agency provide additional information on preferred approaches to applying CDISC to RWD. |
| **Line 109 (and also Line 147)** | Line 109 states that, except for waivers, Sponsors must submit clinical and nonclinical studies in the format under the Study Data Guidance. However, line 147 states that the Sponsor should document rationale for choosing particular CDISC data elements for RWD and document differences. The two statements appear to contradict. | BIO recommends that the Agency provide more details to clarify expectations.<br><br>BIO also recommends that the Agency clarify situations under which a waiver would be granted. For example, whether there might be a need to integrate RWD with CT data that would factor into the decision to grant a waiver. |
| **129-131** | The draft guidance states, "With adequate documentation of the conformance methods used and their rationale, study data derived from RWD can be transformed to SDTM datasets and submitted to FDA in an applicable drug submission." | BIO recommends that the Agency clarify if all study data derived from RWD *should* be transformed to SDTM. Specifically, the Agency should clarify if it would be acceptable to use the study analysis datasets (ADaM format applied) derived from RWD directly without SDTM mapping.<br><br>BIO recommends that the Agency provide additional clarity on the scope of expectations around standards. For example, will Technical Rejection Criteria be applicable to RWD at this point? |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| ***D. Considerations for Mapping Real-World Data to Study Data Submission Standards*** | | |
| **Entire Section** | The draft guidance does not provide any specific mapping tools.<br><br>The draft guidance acknowledges that there may be concepts/terminology that may not be possible to map precisely to SDTM standards and that Sponsors should describe the challenges faced in mapping in the Study Data Reviewer's Guide. Indeed, for RWD studies, it may not always be possible to meet the same level of SDTM compliance as is currently expected for clinical studies, even after best efforts are applied for mapping.<br><br>Additionally, some fields may exist in free text format, such as disease histology, etc. These present additional challenges in mapping, as there is an absence of a standardized code list. | There are a finite number of RWD data sources, especially when you consider the sources most commonly used. BIO recommends that the Agency provide or encourage the collaborative development of crosswalks or mapping from some of these (e.g., from OMOP CDM data to CDISC SDTM). Further, an example of such mapping could be included as an appendix in the final guidance.<br><br>BIO recommends that the Agency comment on its expectations around the level of SDTM compliance and flexibility for RWD datasets.<br><br>BIO recommends the Agency work closely with key stakeholder groups, including data standards consortiums, data providers, industry, and other health authorities, to help advance harmonized standards and coding for RWD. |
| **Lines 143-145** | The draft guidance states, **"**In such cases, Sponsors should document the potential impact of mapping the sex variable or other variables to CDISC's terminology on the study findings." | BIO recommends that the Agency clarify where in the submission this information should be documented. Specifically, the Agency should clarify if the study data reviewers guide is the correct place to document this information, or does this information need to be presented in earlier documentation (e.g., study protocol). |
| **Lines 148-150** | "The sponsor should provide a description of the general approach and anticipated impact of data mapping as a part of or in an appendix to the ***Study Data Reviewer's Guide*** to highlight the domains involved. | BIO recommends that the Agency provide a more detailed appendix with more details on the derivation. |
| **Line 148** | Mapping of RWD sources to CDISC data elements and generating the appropriate documentation to | Similar to our comments above, BIO recommends that the Agency recognize that there would be efficiencies gained if |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| | explain the mapping will require time and resources of Sponsors and could become inefficient. | other common data models are acceptable to submit to FDA for certain types of RWD submissions. |
| **Line 151** | There is likely to be high variability in format and content of a data dictionary, as well as placement on the eCTD backbone, without further guidance from the Agency. | BIO recommends that the Agency provide more detail on the expectations for a data dictionary, e.g., structure, format, location on eCTD backbone. |
| **Line 151** | If original values are maintained in the data, the original meaning is not lost through transformation and is available for further use of the data. | BIO recommends that the Agency consider, where it is appropriate, to include the ”raw “collected values as supplemental qualifiers in the SDTM to aid transparency and maintain data integrity for future use. |
| **Lines 151-153** | The draft guidance states, “Furthermore, the sponsor should include a data dictionary that documents the definition of every data element used and all relevant information about the element, such as its relationships to other data, origin, usage, and format.” | BIO recommends that the Agency clarify as to whether it would be acceptable to only provide an ADaM data submission package, with the link of the RWD under the same Define.XML for the instance where RWD cannot be transformed to SDTM datasets. |
| ***E. Considerations for Data Transformations*** | | |
| **Entire Section** | | FDA is encouraged to collaborate with other data standards entities that, in addition to the data standards-setting bodies mentioned throughout the draft guidance, also establish data standards for RWD used in research. For example: the ONC's USCDI (https://www.healthit.gov/cures/sites/default/files/cures/2020-03/USCDI.pdf), the FHIR accelerator project Vulcan (https://www.hl7.org/vulcan/), OMOP and OHDSI (https://<br><br>www.ohdsi.org/data-standardization/the-common-data-model/).The NIH use of Common Data Elements (CDE) (https://grants.nih.gov/grants/guide/notice-files/not-lm-21-005.html),<br><br>These data standards organizations should be critical collaborators with FDA in defining data standards for RWD. |
| **IV. GLOSSARY** | | |

| SECTION | ISSUE | PROPOSED CHANGE |
|---|---|---|
| **Lines 260-264** | The draft guidance states, "Differences in the coding systems used between real-world data (RWD) and traditional clinical trial data can usually be addressed using the Define-XML file, which is included in all standard Study Data Tabulation Model (SDTM) submissions"<br><br>The Alias element in Define.xml is used for the CDISC controlled terminology C-Code, so it already has a purpose. Similarly, Decode has an existing purpose. There is a potential to overload Define.xml elements (using the same elements for different purposes) which may lead to not being able to represent the information when two or more pieces of information are to go into the same element. | BIO recommends that Terminology mapping should be kept separate.<br><br>BIO also recommends that the Agency provide more guidance on how the Sponsor reassigns coding from raw data into SDTM is warranted. |
| **Lines 322-323** | The draft guidance states, "…mapped structured definitions"<br><br>It is not clear what mapped structured definitions means. | BIO recommends that the Agency clarify the meaning behind "mapped structured definitions". |
| **APPENDIX** | | |
| **Entire Section** | The Appendix was particularly helpful, as well as the aforementioned Conformance Guide. One common occurrence that is not addressed, however, are cases where there is not a 1:1 mapping of terminologies. We see this frequently where one ICD-9 code has expanded to multiple ICD-10 codes. An example of a non-1:1 mapping might increase the compliance as preferred by FDA. | BIO recommends that the Agency provide case examples where there is not a 1:1 mapping of terminology |