PhRMA
RESEARCH • PROGRESS • HOPE
950 F STREET, NW, SUITE 300 • WASHINGTON, DC 20004 • 202-835-3400 • PhRMA.org

February 4, 2022

Division of Dockets Management (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Room 1061
Rockville, MD 20852

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

To Whom It May Concern:

The Pharmaceutical Research and Manufacturers of America (PhRMA) is pleased to submit these comments to the Food and Drug Administration (FDA or the Agency) in response to the draft guidance, *Data Standards for Drug and Biological Product Submissions Containing Real-World Data.*[1] PhRMA supports the Agency's ongoing efforts to implement its Real-World Evidence (RWE) Program in accordance with mandates and commitments made under the 21st Century Cures Act (Cures) and the Prescription Drug User Fee Act VI (PDUFA VI), respectively.[2,3] The use of real-world data (RWD) and its analysis to support regulatory decisions regarding the safety and efficacy of drugs and biologics offers the potential to help make drug development more efficient, predictable, and focused on what is most meaningful to patients.

PhRMA is a voluntary, nonprofit association that represents the country's leading biopharmaceutical research and biotechnology companies, which are devoted to discovering and developing medicines that enable patients to live longer, healthier, and more productive lives. Since 2000, PhRMA member companies have invested more than $1 trillion in the search for new treatments and cures, including $91.1 billion in 2020 alone.

I. GENERAL COMMENTS

PhRMA appreciates that the Agency is providing much needed clarity to sponsors and RWD vendors regarding acceptable data standards for regulatory submissions with the publication of this draft guidance. PhRMA acknowledges that applying the current FDA data standards to evolving and emerging data sets is challenging, as many of these data standards were created when the healthcare data ecosystem was not as diversified, heterogeneous, or technologically advanced.[4]

---

[1] FDA, Data Standards for Drug and Biological Product Submissions Containing Real-World Data. Available at: https://www.fda.gov/media/153341/download.
[2] Federal Food, Drug, and Cosmetic Act (FDCA) § 505F.
[3] FDA, PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2018 Through 2022 (PDUFA VI), Section I.6., Available at: https://www.fda.gov/media/99140/download.
[4] Weber et al. (2014). Finding the Missing Link for Big Biomedical Data. Available at: https://jamanetwork.com/journals/jama/fullarticle/1883026.

While the expansion of RWD types and sources poses challenges to data standardization, it also presents an opportunity for the FDA to re-evaluate data standards for existing data sources and collaborate with clinical researchers, data scientists, standards development organizations, bioinformaticians, technologists, other U.S. government agencies, and other health authorities to better organize data standards for current and future RWD use in a globally harmonized manner. We strongly encourage FDA to work with these stakeholders as the Agency establishes modern data standards best suited to the evolving and emerging RWD assets.

In the final guidance, PhRMA recommends that the Agency ensure the consistent application of and reference to other recently published RWD/E guidances as appropriate, including: (1) *Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products*, (2) *Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products*, and (3) *Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products*.[5,6,7]

Given the interdependencies between all of these draft guidances in conducting an RWE research program and in informing FDA's regulatory decision-making, it will be important for the Agency to consider the broader implications across all RWD/E guidances. PhRMA encourages the Agency to continue to consider public comments on this and the other aforementioned RWD/E draft guidances in their totality when finalizing the RWD/E guidance series.

In addition to the general comments above, PhRMA provides the following specific comments.

II. SPECIFIC COMMENTS

A. Considerations for Section 745A(a) of the FD&C Act for the Implementation of RWD Study Data Standards

FDA notes in the draft guidance that "sponsors submitting clinical and nonclinical study data (including those derived from RWD sources) in submissions subject to Section 745A(a) of the FD&C Act are required to use the formats described in the Study Data Guidance and the supported study data standards listed in the Catalog." Given the recognized challenges with

[5] FDA, Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products. Available at: https://www.fda.gov/media/152503/download.

[6] FDA, Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products. Available at: https://www.fda.gov/media/154449/download.

[7] FDA, Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products. Available at: https://www.fda.gov/media/154714/download.

conforming RWD with the standards set forth in the Study Data Guidance and the Study Data Catalog (Catalog), as well as FDA's stated "plans to issue further guidance and/or update the Catalog with standards for study data that are derived from RWD sources," PhRMA encourages FDA to exercise flexibility in granting waivers to 745A(a) for RWD, particularly for those studies that were initiated prior to issuance of this draft guidance where sponsors did not have the benefit of clarity that FDA has provided here and in other recently published RWD/E draft guidances.

B. Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products and Specific Data Exchange Format

PhRMA appreciates FDA's acknowledgment of the challenges involved in standardizing study data derived from RWD sources for inclusion in applicable drug submissions (lines 85-93). PhRMA encourages the Agency to include best practices to remediate challenges in RWD standardization to better support the use of RWD in regulatory decision-making.

PhRMA refers the Agency to our comments on FDA's draft guidance on *Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products* for additional remarks on RWD standardization and harmonization between sources.[8]

C. Data Access for Studies Initiated Prior to this Draft Guidance

RWD is sourced from a variety of different data partners, and sponsors generally must negotiate with these data partners for access and use of the data. PhRMA appreciates FDA's guidance here, and in other recently published draft guidances on RWD/E, clarifying the Agency's expectations for the type of data that should be included in submissions to FDA. These expectations will help inform future work with data partners when generating RWD for purposes of regulatory decision-making. However, for studies generating RWD initiated prior to the publication of this (and related) draft guidances, sponsors may not necessarily have access to the data FDA has now recommended be included in submissions to FDA. Accordingly, PhRMA encourages FDA to maintain flexibility in accepting RWD-derived datasets from sponsors for these studies. Moreover, it is unclear that sponsors will be able to secure the information that FDA seeks going forward, and PhRMA recommends that the Agency acknowledge that sponsors should discuss these challenges with the relevant review division during development.

Even for studies initiated after publication of this draft guidance, there will be challenges in conforming RWD to the data standards currently supported in the Catalog. For example, where a sponsor uses multiple sources of RWD in a regulatory submission, there will likely be challenges in conforming multiple sources into a single standard, even when utilizing the recommendations FDA outlines in this draft guidance. PhRMA encourages FDA to provide its

[8] See Appendix.

current thinking on how sponsors may address standardization challenges in this circumstance. PhRMA recommends that the guidance acknowledge that sponsors may be working with various data partners who generate RWD in different formats, and that challenges/mitigations related to information access and standardization may exist but could be part of a sponsor's discussions with the Agency.

D. Current Data Standards Catalog and Future Updates

PhRMA acknowledges the Agency is planning to issue further guidance and/or update the Catalog with standards for study data that are derived from RWD sources. As FDA updates the Catalog to establish data standards to best facilitate the adoption and use of RWD in regulatory submissions, PhRMA encourages the Agency to recognize that implementation of practices to conform to any newly identified data standards may take time as sponsors conduct impact assessments and make appropriate process changes. A public and transparent process to updating the Catalog will best help enable sponsors and other stakeholders to conform with any newly added data standards. PhRMA appreciates FDA's stated intent to issue further guidance and/or update the Catalog and offers four recommendations to facilitate the use of RWD in regulatory submissions.

First, PhRMA recommends that FDA address healthcare-related data standards, terminologies, formats, and methodologies in future updates and guidance. The use of RWD for RWE encompasses a set of broad, complex, and highly variable data sources, as FDA rightly acknowledges in the draft guidance. To facilitate use of RWD, PhRMA recommends that the Agency update the Catalog to support healthcare-related terminologies (e.g., ICD-10, RxNorm, etc.), which may not currently align to Clinical Data Interchange Standards Consortium (CDISC) Controlled Terminology.

Second, in looking to update the Catalog, PhRMA encourages FDA to assess and evaluate how the current data standards conform with RWD used in studies such as pragmatic clinical trials, external control studies, long-term follow-up studies, and prospective and retrospective non-interventional studies, as well as how they conform with RWD from different data sources like EHRs, medical claims, and registries.[9] When considering appropriate data standards for RWD, Analysis Data Model (ADaM)-like principles could be applied and modified to appropriately capture RWD. For example, PHUSE EU recently performed an assessment of challenges and various solutions for programmers when dealing with RWD non-interventional studies.[10] This type of assessment and learnings from the PHUSE EU paper could be leveraged when considering appropriate data standards.

Third, PhRMA recommends the Agency seek stakeholder feedback in updating the Catalog with additional standards for data using RWD sources, consider other common data models

[9] Andre, et. al. (2019). Trial designs using real-world data: The changing landscape of the regulatory approval process. Available at: https://onlinelibrary.wiley.com/doi/full/10.1002/pds.4932.
[10] Phuse Data Standards for Non-Interventional Studies. Available at: https://phuse.s3.eu-central-1.amazonaws.com/Deliverables/Optimizing+the+Use+of+Data+Standards/Data+Standards+for+Non-interventional+Studies.pdf.

such as the Sentinel CDM and the Observational Health Data Sciences and Informatics (ODHSI) Observational Medical Outcomes Partnership (OMOP), and leverage existing standards.[11] Specifically, PhRMA recommends FDA reference *ISO 11615*[12] to ensure consistency in the terminology and controlled vocabularies used to allow drug comparison. Further, PhRMA encourages FDA to collaborate with other data standards entities and other ongoing data standards efforts within the U.S. government, in addition to the data standards-setting bodies mentioned throughout the guidance, that also establish data standards for RWD and to share lessons learned with stakeholders in a public setting.[13] PhRMA notes that global harmonization on data standards is critical and recommends the Agency work with other regulatory agencies to seek harmonization on RWD standards.

Fourth, PhRMA recommends the Agency describe a process for contacting the Agency if there is a suggested addition to the Catalog.

E. Data Mapping and Transformation

FDA appropriately acknowledges the challenges in mapping and transforming RWD into data that meet current FDA-supported data standards. The draft guidance notes that there may be concepts/terminologies that may not be possible to map precisely to CDISC Study Data Tabulation Model (SDTM) standards and that sponsors should describe the challenges faced in mapping in the Study Data Reviewer's Guide. For RWD studies, it may not always be possible to format RWD precisely in CDISC SDTM standards as is currently expected for clinical studies, even after best efforts are applied to mapping. PhRMA recommends including "raw" collected values as supplemental qualifiers in the SDTM to aid transparency and maintain the integrity of data for future use. By maintaining original values in the data, the original meaning is not lost through transformation and is available for further use of the data. As this may not be feasible in all cases, PhRMA recommends that the guidance provide the Agency's current thinking on its expectations in circumstances where sponsors cannot fully meet CDISC SDTM standards for RWD datasets. Additionally, PhRMA recommends FDA include an example in the appendix of how Required/Expected CDISC variables that are not available in the RWD source should be addressed.

In the final guidance, PhRMA encourages the Agency to reference the recently published FHIR to CDISC Joint Mapping Implementation Guide by CDISC and acknowledge that

---

[11] Sentinel Common Data Model (https://www.sentinelinitiative.org/methods-data-tools/sentinel-common-data-model); OHDSI's OMOP (https://www.ohdsi.org/data-standardization/the-common-data-model/).

[12] International Organization for Standardization, Health informatics — Identification of medicinal products — Data elements and structures for the unique identification and exchange of regulated medicinal product information. Available at: https://www.iso.org/standard/70150.html.

[13] 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/); Common Data Models Harmonization FHIR Implementation Guide (http://hl7.org/fhir/us/cdmh/2019May/); OHDSI's OMOP (https://www.ohdsi.org/data-standardization/the-common-data-model/), NIH use of Common Data Elements (CDE) (https://grants.nih.gov/grants/guide/notice-files/not-lm-21-005.html); NLM Health IT and Health Data Standards (https://www.nlm.nih.gov/healthit/index.html); CMS Health Informatics and Interoperability Group (https://www.cms.gov/regulations-and-guidance/guidance/interoperability/index).

sponsors may follow the recommendations therein to remap data from EHR to CDISC SDTM variables.[14] PhRMA recommends that the Agency initiate user acceptance testing/pilot testing with sponsors to gain better understanding of mapping RWD sources to SDTM and share lessons learned, as appropriate, per current PDUFA VI and PDUFA VII goals.[15] Additionally, an example of EHR to CDISC SDTM mapping could be included as an example in the appendix of the final guidance and the Agency could encourage the collaborative development of common data models that are acceptable to submit to FDA for certain types of RWD submissions.

PhRMA requests additional guidance related to the issues and challenges raised when multiple RWD sources are used and/or linked, including recommendations on how to address divergences among multiple data sources and FDA-supported data standards. Generally, there is an acknowledgment that as data sources increase, the terminology for various concepts (e.g., gender identity) increases. PhRMA recommends FDA provide detailed guidance on how to map data standards and terminologies, including data sources and processes that may have used non-structured data to generate data elements (e.g., natural language processing).

Another issue that often arises when generating RWD derived from EHRs is inconsistent dating of items. For example, medication start dates are more reliably complete than medication end dates. Some systems will automatically pull forward data into subsequent visits unless deliberately overwritten, and there may or may not be a flag for such "auto-populated" data. PhRMA recommends additional information in the final guidance on best practices to address these circumstances.

Finally, PhRMA requests clarification on documenting the impact of data mapping and transformation on the source data. 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. PhRMA encourages FDA to work closely with key stakeholder groups, including data standards consortia, data providers, industry, other U.S. government agencies, and other health authorities, to help advance harmonized standards and coding for RWD.

F. Documentation of Processes & Conforming to Currently Supported Standards for RWD

PhRMA agrees with FDA that adequate documentation during RWD curation and transformation is critical. However, the draft guidance is limited regarding the documentation needed during data curation and data transformation steps (lines 98-101).

---

[14] FHIR to CDISC Joint Mapping Implementation Guide v1.0. Available at: https://www.cdisc.org/standards/real-world-data/fhir-cdisc-joint-mapping-implementation-guide-v1-0.

[15] FDA, PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2018 through 2022, Section IV. Available at: https://www.fda.gov/media/99140/download; FDA, PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2023 through 2027, Section IV. Available at: https://www.fda.gov/media/151712/download.

To help sponsors understand FDA's recommendations, PhRMA requests the Agency define the term "metadata-driven audit trail" as used in the draft guidance.

PhRMA also requests the Agency clarify expectations for an exemption or waiver related to challenges in formatting RWD into an accepted standard listed in the Catalog. The Agency should expand further or provide example use cases regarding this statement.

G. Location and Further Expansion of Recommended Documentation

PhRMA recommends the Agency expand upon its recommendation that sponsors "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." An example of this recommended "description" would help sponsors understand the level of data FDA expects.

To help facilitate organization of regulatory submissions, PhRMA recommends that FDA be explicit in describing where certain information should be included. 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. FDA also recommends that sponsors include a "Data Dictionary"[16] in their submission. As RWE studies have previously been included in Module 5, under m5-3-5-4-other-study-reports, PhRMA recommends RWD datasets reside in the same section as the report, structured according to the current FDA-supported standards (i.e., SDTM/ADaM and Define.xml). PhRMA also suggests FDA align its recommendations with the USCDI v2 as these standards for interoperability (linking and sharing) should be part of RWD data standards used for RWE in regulatory decision-making.[17]

Finally, PhRMA notes the Define-XML (all existing versions) was not developed with RWD in mind. For example, CDISC will require enhancements to Define-XML v1.1 (e.g., Alias element is not a suitable option in all cases) to support greater flexibility for RWD requirements. There is likely to be high variability in format and content as well as placement in the eCTD backbone without further guidance from the Agency.

III. CONCLUSION

PhRMA appreciates FDA's issuance of this draft guidance and looks forward to future engagement with FDA as the Agency continues to implement its RWE Program. RWD/RWE will play a key role in the future of innovative drug development by ensuring that important therapies reach patients in a timely and efficient manner.

---

[16] Such a dictionary, in practice, is often specified in the Define-XML file. PhRMA requests the Agency include clear expectations for the Data Dictionary (i.e., the structure, format, and location in the eCTD backbone, including whether the data dictionary must be an additional document or if it can be incorporated within the Define.xml).

[17] USCDI Version2. Available at: https://www.healthit.gov/isa/united-states-core-data-interoperability-uscdi#uscdi-v2).

Respectfully submitted,

/s/	/s/
Kristin Dolinski Deputy Vice President, Science and Regulatory Advocacy	Rebecca Nebel, PhD Director, Science and Regulatory Advocacy

Appendix

PhRMA Comments on *Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products*.

PhRMA
RESEARCH • PROGRESS • HOPE
950 F STREET, NW, SUITE 300 • WASHINGTON, DC 20004 • 202-835-3400 • PhRMA.org

January 24, 2022

Division of Dockets Management (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Room 1061
Rockville, MD 20852

RE: Docket No. FDA-2020-D-2307: Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products.

To Whom It May Concern:

The Pharmaceutical Research and Manufacturers of America (PhRMA) are pleased to submit these comments to the Food and Drug Administration (FDA or the Agency) in response to the draft guidance, *Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products* (draft guidance).[1] PhRMA supports the Agency's ongoing efforts to implement its Real-world Evidence (RWE) Program in accordance with mandates and commitments made under the 21st Century Cures Act and the Prescription Drug User Fee Act VI, respectively.[2,3] The use of real-world data (RWD) and its analysis to support regulatory decisions regarding the safety and efficacy of drugs and biologics offers the potential to make drug development more efficient and productive.

PhRMA is a voluntary, nonprofit association that represents the country's leading biopharmaceutical research and biotechnology companies, which are devoted to discovering and developing medicines that enable patients to live longer, healthier, and more productive lives. Since 2000, PhRMA member companies have invested more than $1 trillion in the search for new treatments and cures, including $91.1 billion in 2020 alone.

I. GENERAL COMMENTS

PhRMA commends the Agency for releasing draft guidance that provides key principles on the reliability and use of relevant data from electronic health records (EHR) and medical claims data (claims data) to be used for regulatory decision-making.

[1] 86 FR 54219; Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry (September 2021). Available at: https://www.fda.gov/media/152503/download.

[2] Federal Food, Drug, and Cosmetic Act (FDCA) § 505F.

[3] FDA, PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2018 Through 2022 (PDUFA VI). Section I.6. Available at https://www.fda.gov/media/99140/download.

PhRMA also commends the Agency for underscoring the principles of transparency in conducting RWE studies including pre-specification and submission of protocols and statistical analysis plans to the Agency. The draft guidance establishes best practices that the entire research community can utilize, including biopharmaceutical companies, data firms, regulators, and academics. The draft guidance takes meaningful steps to define what constitutes fit-for-purpose RWD and research practices with an emphasis on quality, standardization, and provenance. PhRMA notes the practical application of the draft guidance when used in combination with the 2013 FDA guidance, Use *of Electronic Health Record Data in Clinical Investigations,* and refers to our previous comments on the subject.[4] PhRMA believes that this draft guidance represents an important step forward for the Agency and will inform the Advancing RWE Program pilot outlined in PDUFA VII.[5]

The draft guidance publicly details FDA's recommendations on a number of RWD issues for the first time, including on the types of data that sponsors should include in submissions and/or ensure FDA has access to. Recognizing that data may be generated and owned by many different parties, PhRMA notes that, for studies initiated prior to publication of this draft guidance, sponsors may not have access to the data to generate the recommended documentation in submissions to FDA. Accordingly, PhRMA requests that FDA acknowledge these challenges in the draft guidance, exercise flexibility, and engage in appropriate knowledge sharing in this space, as sponsors work collaboratively with other partners to update their internal processes and practices for engagement to reflect the recommendations in this draft guidance.

Further, PhRMA is interested in collaborating further with the Agency and other relevant stakeholders to develop a framework to recommend quantitative and qualitative metrics of data quality that are adequate to ensure the completeness, accuracy, and plausibility of the relevant data source. We believe that more specific guidance is needed to help enable sponsors and data providers to generate consistent assessment reports for data quality that will meet the Agency's needs. PhRMA proposes that industry and data providers could work with the Agency to develop such a framework. FDA could then include the framework of data quality metrics in an appendix to the final guidance that outlines key elements to include in regulatory submissions.

A. Scope of Guidance

As the draft guidance explicitly focuses on EHRs and medical claims data, PhRMA recommends that FDA revise the draft guidance to also cover "pharmacy/prescription claims data" in addition to "medical claims data." Pharmacy/prescription claims provide valuable information for purposes of evaluating drug safety and efficacy and are an essential part of insurance claims data that are commonly used by sponsors in RWD studies.

---

[4] See PhRMA comments on FDA Draft Guidance for Industry - Use of Electronic Health Record Data in Clinical Investigations. Available at https://downloads.regulations.gov/FDA-2016-D-1224-0026/attachment_1.pdf.

[5] FDA, PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2023 Through 2027 (PDUFA VII). Available at https://www.fda.gov/media/151712/download.

As currently written, the draft guidance appears to only cover claims data generated by inpatient and outpatient services covered under medical benefits.

PhRMA recommends that FDA reference approaches such as the target trial framework[6] and estimand thinking process as described in the ICH E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials [7] as these can be useful in formalizing the study question when this concerns the causal effect of a drug on an efficacy or safety variable.

The current draft guidance indicates that study design and analysis will be covered in a future draft guidance.[8] PhRMA appreciated this statement of intent and encourages the Agency to be as expeditious as possible in publishing this draft guidance to facilitate the uptake of RWE.

B. Consistent Application and Interdependencies of Additional RWE Guidances

PhRMA acknowledges that draft guidance on 1) *Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products,*[9] *2) Data Standards for Drug and Biological Product Submissions Containing Real-World Data,*[10, 11] and 3) *Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products*[12] have been published by the Agency for comment. Additionally, PhRMA understands that FDA plans to issue draft guidance on study designs and analysis.[13] Given the interdependencies between all of these guidances in conducting an RWE research program and meeting rigorous standards for

---

[6] Hernan MA, Robins JM, Am J Epidemiology (2016) 183:758.

[7] FDA/International Council for Harmonisation, Guidance for Industry - E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. Available at https://www.fda.gov/media/148473/download.

[8] Line 110, 86 FR 54219; Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry (September 2021).

[9] 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 (December 2021). Available at https://www.fda.gov/media/154714/download.

[10] FDA, Draft Guidance for Industry – Data Standards for Drug and Biological Product Submissions Containing Real-World Data (October 2021). Available at https://www.fda.gov/media/153341/download.

[11] See comments filed by PhRMA on Feb. 8, 2022, in response to Draft Guidance for Industry – Data Standards for Drug and Biological Product Submissions Containing Real-World Data (October 2021).

[12] FDA, CDER Guidance Agenda, New & Revised Draft Guidance Documents Planned for Publication in Calendar Year 2021 (July 2021). Available at https://www.fda.gov/media/134778/download.

[13] Line 110, 86 FR 54219; Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry (September 2021).

regulatory decision-making, PhRMA encourages the Agency to continue to consider public comments on this and the other draft guidances after the close of their respective comment periods to allow stakeholders the ability to comment on broader implications across all guidances in this RWE series.

In addition to the general comments above, PhRMA provides the following specific comments as well as line-by-line comments to ensure consistent and coordinated approach to terminology in the appendix.

II. SPECIFIC COMMENTS

A. Validation

PhRMA requests that FDA clarify expectations, increase flexibility through a fit-for-purpose approach, and better outline recommended steps for validation while acknowledging that validation may be iterative.

While parts of the draft guidance acknowledge that the need for extensive validation of eligibility criteria, exposures, outcomes, and covariates depends on multiple factors (expected treatment effect size, scientific question of interest, previous validation studies, etc.), other parts of the draft guidance do not appear to allow for this flexibility. For example (Line 820): "FDA expects validation of the outcome variable to minimize outcome misclassification." Additionally, section headers (*Validation of Exposure* [Line 666]; *Validation of Outcome* [Line 818]; *Validation of Confounders and Effect Modifiers* [Line 1066]) imply that formal validation studies for relevant variables will be the default recommendation. It would be helpful for the guidance to be clear and consistent in referencing that the need for and extent of validation required for a given study will depend on the specific study question and regulatory question at-hand. Variables of importance could be pre-specified for validation and sponsors may submit these plans to FDA in the protocol for discussion and to obtain feedback on before the study begins. Initial discussions with FDA could include feasibility (e.g., applicable requirements around privacy and data anonymization may prohibit validation against medical records) and validation study plans. Validation of all operationalized covariates may not be the most efficient means of addressing potential data quality issues. PhRMA suggests that FDA state in the guidance that important covariates that may have more complex operational definitions be prioritized in the validation plan.

The draft guidance emphasizes validation for confounders and assumes the ability to collect all potential confounders and covariates. Given that some of these data elements will not be captured consistently in routine care, we ask FDA to take a risk-based approach that balances scientific rigor with the pragmatic realities of real-world data collection today. One way to achieve this would be to consider which potential confounders are likely to have the greatest impact based on the research question, clinical and regulatory context including the therapeutic area and study population. Early bias analysis could be used to examine the range of potential impacts and plausibility during the feasibility phase. PhRMA understands that these points may be included in the planned draft guidance on study design and

analytical methodologies. If so, PhRMA recommends that FDA ensure the relevant sections are cross-referenced when the guidances are finalized.

Where there may be challenges in validation against the best available benchmark or where such a reference standard does not exist, PhRMA asks the FDA to consider providing examples of acceptable approaches including the use of analytic techniques, such as component analyses or quantitative bias analysis, as a means of quantifying potential impact on study results of uncertainty around classification of key variables. PhRMA encourages FDA to take a flexible approach recognizing study-specific context and the variation in available data sources which may require trade-offs. PhRMA recommends including any particular considerations that may be unique in validating RWD that are being used in conjunction or in comparison to clinical trial data such as with a single arm trial and an external RWD control.

PhRMA suggests that FDA clarify that a sponsor may prespecify what aspects of validation may be conducted before protocol finalization and also how the validation plan covers the on-going study. PhRMA recommends a stepwise approach to a validation plan that starts with more general data characterization, before developing a final validation plan for key variables and that an analysis plan (e.g., sensitivity analyses) might be based on information ascertained from the on-going study.

Lastly, with regard to disease areas such as rare diseases where there may be more limited RWD available to inform benefit-risk evidence generation, PhRMA encourages FDA to embrace innovative approaches and apply flexibility when evaluating the approach to validation and assessment of these datasets.

B. Documentation

The draft guidance includes numerous recommendations around the inclusion of certain information in the "Study Protocol", "Analysis Plan", "Study Report", and/or "Study Documents", as well as recommendations to "describe," "document," and/or "assess" information without a particular document referenced, but it is not clear as to where this information should be included. Accordingly, PhRMA requests the following clarifications:

1)Information FDA expects to be in materials that should be pre-specified and submitted to the Agency prior to conducting the study (i.e., Protocol and Analysis Plan);
2) Information that could go into materials that would only be provided to FDA after study completion (i.e., Study Report); and
3) Information that only needs to be available upon request in case of an inspection or audit by the Agency (e.g., data quality reports).

PhRMA notes that these recommendations are interspersed throughout the full length of the draft guidance. It would be helpful if the draft guidance included an appendix, or that a question-and-answer companion document be generated to specifically summarize/extract all these recommendations for what should go into a protocol, statistical analysis plan (SAP), study report, etc.

C. Data Access

Biopharmaceutical companies often partner with data vendors, insurance providers, registry holders, and other entities that own the source RWD used to generate RWE. PhRMA recommends that the guidance acknowledge that sponsors may be working with various partners that own this information and that challenges/mitigations related to information access may exist and could be part of the sponsor's discussions with the review division. For studies initiated prior to publication of this draft guidance in particular, sponsors were not aware of FDA's recommendations for including certain data in regulatory submissions or otherwise ensuring FDA access to certain data. Accordingly, PhRMA recommends that FDA recognize these challenges faced by sponsors of such studies and confirm that the Agency will exercise flexibility in accepting certain data in regulatory submissions.

PhRMA requests that FDA clarify the circumstances where patient-level versus aggregated RWD may support regulatory decisions as part of the totality of evidence. Generally, PhRMA agrees that the FDA's recommendation for patient-level data to be included in regulatory submissions is appropriate for interventional studies provided that these data are providing pivotal evidence to the regulatory decision. Regarding non-interventional studies, the need for access to patient-level data should be discussed on a case-by-case basis with the appropriate regulatory review division.[14] Accordingly, we recommend that FDA recognize in the draft guidance that, in certain circumstances, patient-level data may not be accessible, and in such cases, sponsors should consult with the relevant review division on other acceptable approaches.

Finally, PhRMA requests clarity on the distinction between information expected for evaluation during the protocol development stage versus during the study conduct and final report.

D. Assessment of Data Source Relevance

PhRMA requests that FDA clarify whether the provision of background information on a data source is expected at a general level or should be specific to the study cohort and variables of interest. PhRMA is concerned that the current view outlined in the draft guidance appears to suggest that this should be done on a study-by-study basis.

PhRMA requests that FDA clarify what information about prior experience with data from a specific RWD source may be relied upon for other submissions (when the sponsor has access to the data to be relied upon). PhRMA recommends that FDA consider a streamlined approach akin to the approach adopted by HMA/EMA that establishes a 'meta-data' resource[15] where, in appropriate circumstances, general characteristics of the patients in a

[14] While PhRMA recognizes that this may be a case-by-case basis, guidance on overarching considerations that the Agency would take into account would be informative.

[15] EMA, Technical workshop on real-world metadata for regulatory purposes. Available at https://www.ema.europa.eu/en/documents/other/summary-report-technical-workshop-real-world-metadata-regulatory-purposes_en.pdf.

database, the standard of care that they receive, and variables such as endpoints be extrapolated from one study to another, and FDA would not expect a sponsor to re-validate. For example, if a sponsor established and consistently applied a method for validating mortality in a specific database and the method were published and/or used in an approval; FDA should confirm in the guidance that the Agency would not expect a sponsor to repeat the validation.

PhRMA requests that FDA describe what specific information about generalizability of study results from non-US data sources should be submitted to FDA that are unique from what information should be provided for other cohorts and sub-populations derived from RWD. PhRMA requests that FDA clarify the circumstances or provide examples of situations where certain factors might affect the generalizability of study results to the US population.

E. FDA Meetings and Engagement

PhRMA believes that it is important for sponsors to have the opportunity to meet with FDA early in the RWE study design process to discuss RWE study objectives and protocols with relevant FDA review divisions. PhRMA acknowledges that the current and forthcoming draft guidances would provide a level of regulatory clarity that could serve to streamline Agency-sponsor interactions. However, we are concerned that the type of protocol and SAP submission process envisioned under the draft guidance could take several months-to-years to develop per research project.

PhRMA acknowledges that FDA addresses how sponsors should engage with the Agency in early stages of designing a non-interventional study intended to support a marketing application in the draft guidance, *Considerations for the Use of Real-World Data and Real-World Evidence to Support Regulatory Decision-Making for Drug and Biological Products*.[16] PhRMA requests that FDA cross reference this information in the final guidance for clarity. A shared understanding of the details and expectations for sponsor interactions with the review divisions including process, the minimum amount of information sponsors should be prepared to provide to support a robust discussion, type(s) of meeting, and suggested timing/timelines will be helpful.

PhRMA believes that it would be helpful to identify streamlined and efficient FDA-sponsor communication methods to facilitate rapid agreement about what a study program entails that will allow for evidence generation to support regulatory decision-making. PhRMA looks forward to the FDA's initiation of the Advanced RWE Pilot Program and new Type D meeting under PDUFA VII,[17] which is meant to streamline this type of scientific dialogue, enhance FDA's preparation of RWD/E guidances and incorporate learnings from case-studies. PhRMA anticipates that this should facilitate finding productive solutions and sharing those publicly

[16] 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 (December 2021). Available at https://www.fda.gov/media/154714/download.

[17] FDA, PDUFA Reauthorization Performance Goals and Procedures Fiscal Years 2023 Through 2027 (PDUFA VII). Available at https://www.fda.gov/media/151712/download.

with stakeholders may alleviate many challenging issues without necessitating multiple formal meetings.

PhRMA agrees with FDA's encouragement for sponsors to seek feedback from or have discussions with the Agency on the proposed EHR or medical claims study (Line 100), including the need for variable validation (Line 469), the ability of the proposed data source to capture potential temporal changes and any impact on the study's internal validity (Line 534), and the proposed study outcome definitions (Line 771).[18]

The draft guidance specifies that the "relevant FDA review division" should be engaged in providing feedback or participating in discussions. PhRMA strongly urges the Agency to ensure participation from other relevant offices with RWE expertise (e.g., Office of Medical Policy (OMP), Office of Surveillance and Epidemiology (OSE)), as needed, to support a robust conversation. In addition, it would be helpful to understand the anticipated role of OMP's RWE Subcommittee in FDA-Sponsor meetings on RWE proposals going forward. Articulation of the Subcommittee's role could be particularly helpful in fostering consistency in feedback and expectations across review divisions. PhRMA suggests that the OMP RWE Subcommittee act as an independent review resource that can bring appropriate expertise to all submissions involving RWD/E across review divisions.

F. Case Examples

PhRMA acknowledges that some conceptual examples are described in the draft guidance to illustrate considerations/trade-offs, such as the example for neural tube defects in infants. However, it would be helpful to include additional examples of how EHRs and medical claims data were successfully and unsuccessfully used for efficacy/effectiveness studies. While the draft guidance provides useful information on selection of data sources, for instance, it does not provide recommendations on how submitted data from EHRs and medical claims data have been or could be used by FDA in regulatory decision-making. The guidance would benefit from a broader set of examples representing a spectrum of disease areas, including oncology and rare diseases, where RWE has been used successfully for regulatory decision-making, as well as a broader set of data sources including novel methods of data capture.

[18] 86 FR 54219; Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry (September 2021). Available at https://www.fda.gov/media/152503/download.

Additionally, PhRMA recommends that FDA reference key scientific papers and authoritative perspectives (e.g., ISPOR, ISPE, ENCePP, others[19, 20, 21, 22, 23]) that discuss the use of EHRs/claims data and how to overcome limitations, as well as the highlighted conceptual examples.

G. Acceptable Data Standards

PhRMA refers the Agency to our comments evaluating FDA's draft guidance on *Data Standards for Drug and Biological Product Submissions Containing Real-world Data*[24,25] for additional remarks on data standards for RWD.

1. Common Data Standards

The draft guidance recommends the use of Common Data Models (CDMs) to harmonize the data sources. PhRMA encourages flexibility in the use of different CDMs based on the research questions and data type however, additional Agency clarification on acceptable CDMs beyond the FDA Data Standards Catalog would be beneficial. The research community would benefit from further details on CDMs that are already well-developed, such as OMOP/ODHSI and HL7/FHIR.

2. Data Quality

PhRMA agrees with the recommendations made throughout the draft guidance related to data quality during the accrual, curation, and transformation process (specifically Section VI). PhRMA recommends that FDA clearly define expectations for what information from sponsors should be furnished and when. For instance, the guidance does not provide specific recommendations on documentation needed for 'traceability' or 'provenance.'

---

[19] Dreyer NA, et al. Use of the Grace Checklist for Rating the Quality of Observational Comparative Effectiveness Research. Value in Health, Volume 17, Issue 7, A732. Available at https://www.ispor.org/publications/journals/value-in-health/abstract/Volume-17--Issue-7/Use-of-the-Grace-Checklist-for-Rating-the-Quality-of-Observational-Comparative-Effectiveness-Research.

[20] Berger ML, et al. Good practices for real-world data studies of treatment and/or comparative effectiveness: Recommendations from the joint ISPOR-ISPE Special Task Force on real-world evidence in health care decision making. Pharmacoepidemiol Drug Saf. 2017 Sep; 26(9):1033-1039. Available at https://pubmed.ncbi.nlm.nih.gov/28913966/.

[21] Hall GC, et al. Guidelines for good database selection and use in pharmacoepidemiology research. Pharmacoepidemiol Drug Saf. 2012 Nov; 21: 1-10. Available at https://onlinelibrary.wiley.com/doi/10.1002/pds.2229.

[22] Duke-Margolis Center for Health Policy, Characterizing RWD Quality and Relevancy for Regulatory Purposes (2018). Available at: https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf.

[23] ISPE, Guidelines for Good Pharmacoepidemiology Practices (GPP). 2015. Available at https://www.pharmacoepi.org/resources/policies/guidelines-08027/.

[24] FDA, Draft Guidance for Industry – Data Standards for Drug and Biological Product Submissions Containing Real-World Data (October 2021). Available at https://www.fda.gov/media/153341/download.

[25] See comments filed by PhRMA on Feb. 8, 2022, in response to Draft Guidance for Industry – Data Standards for Drug and Biological Product Submissions Containing Real-World Data (October 2021).

PhRMA recommends that FDA include its thinking on how sponsors can adequately specify data provenance for a dataset from an event prompt to data being in a research-ready format and then to tables, figures, and listings/study report from a sponsor. In this case, PhRMA recommends that FDA confirm that a sponsor should provide FDA with access to data provider SOPs (e.g., for QA/QC) and data management plans plus a protocol/SAP, computer programs, and study report.

H. Medication Adherence

PhRMA suggests that FDA state in the final guidance that medication dosing and supply imputation approaches are important considerations in the design of a RWD study. FDA states that the current draft guidance does not address issues related to medication adherence.[26] PhRMA recommends that this should be considered when FDA addresses study design and statistical methods in future guidance.

I. Ensuring a Consistent and Coordinated Approach on Terminology

In addition to the comments provided above, PhRMA provides the following specific comments in Appendix I in an effort to align on terminologies, expand the Glossary section to be more comprehensive, and to ensure the draft guidance clearly communicates FDA's expectations and facilitates effective sponsor interactions with FDA to aid with efficient and innovative drug development programs (see Appendix I).

III. CONCLUSION

PhRMA appreciates FDA's efforts to develop and provide this draft guidance to facilitate broader use of RWD and RWE, which in turn can facilitate patients' timely access to more innovative medicines. PhRMA looks forward to future engagement with FDA as the Agency continues to develop additional guidance(s), hold stakeholder workshops to share knowledge, and implement an effective RWE Program to support regulatory decision-making. PhRMA believes the groundwork that is being laid with the draft guidance will help ensure that innovative drug development may include RWD/RWE for regulatory decision making and help innovative therapies to reach patients in as timely and efficient manner as possible.

[26] Line 647, footnote 13; 86 FR 54219; Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry (September 2021). Available at: https://www.fda.gov/media/152503/download.

Respectfully submitted,

/s/	/s/
Kristin Dolinski Deputy Vice President, Science and Regulatory Advocacy	Kelly Falconer Goldberg Vice President, Law/Senior Counsel for Biopharmaceutical Regulation
/s/	/s/
Sarah K. Martin, MS, PhD Director, Science and Regulatory Advocacy	Ryan Kaat Assistant General Counsel

Appendix I. ALIGNING ON TERMINOLOGY

Line No.	Existing Text	Comment	Proposed change (if applicable)
Line 32, Footnote 7	"For the purposes of this guidance, the term *clinical studies* refer to all study designs, including but not limited to, interventional studies where the treatment is assigned by a protocol (e.g., randomized, or single-arm trials, including those that use RWD as an external control arm) and noninterventional studies where treatment is determined in the course of routine clinical care – i.e., observational studies (e.g., case-control or cohort studies). Throughout the guidance, FDA uses the terms *clinical studies*, *studies*, and *study* interchangeably."	PhRMA emphasizes the importance of clear definitions and their consistent use to limit confusion regarding study types. Further clarity regarding the definition of *observational studies* is needed, especially as e-CRF is not discussed as a data source.	PhRMA requests that FDA apply a consistent use of the terms *clinical study/studies* or *study/studies*. In addition, clarification of *observational study* would be useful.
Line 57	"Development and validation of definitions for study design elements (e.g., exposure, outcomes, covariates)"	Population is an important study design element.	Development and validation of definitions for study design elements (e.g., exposure, outcomes, covariates, **population**)
Lines 66-69	"For the purposes of this guidance, the term reliability includes data accuracy, completeness, provenance, and	PhRMA suggests the following language be added to the definition of *relevance*: "…and sufficient follow-up time."	"The term relevance includes availability of key data elements (exposure, outcomes, covariates) and sufficient numbers of

Line No.	Existing Text	Comment	Proposed change (if applicable)
	traceability. The term relevance includes the availability of key data elements (exposure, outcomes, covariates) and sufficient numbers of representative patients for the study. “		representative patients for the study **and sufficient follow-up time**.”
Lines 106-108	“This guidance provides recommendations on selecting data sources to maximize the completeness and accuracy of the data derived from EHRs and medical claims for clinical studies. “		“This guidance provides recommendations on selecting **reliable and relevant** data sources to maximize the ~~completeness and accuracy of the data derived from~~ **potential use of** EHRs and medical claims ~~for~~ **in** clinical studies **to support regulatory decisions**.”
Lines 145 - 150	“EHR data are generated for use in clinical care and may also serve as a basis for billing and for auditing of practice quality measures. Data recorded in an EHR system depend on each health care system’s practices for patient care and the clinical practices of its providers. In addition, data collection is limited to the data captured within an EHR system or network, and may not represent comprehensive care (e.g., care	PhRMA notes that the terms EHR and electronic medical record (EMR) are often used interchangeably. However, it may be useful to define EMR databases and distinguish them from EHR databases for the purposes of this discussion. This distinction can help to describe the data gaps in EHR databases that result from care received outside of a particular health system.	PhRMA proposes defining EMR in the glossary to differentiate between an EHR and EMR databases and stating that the guidance focuses on EHR data. PhRMA suggests adding the following sentence to Line 150: “**Similar to claims, EHR data may not accurately reflect presence, specificity, or severity of a particular disease.**”

Line No.	Existing Text	Comment	Proposed change (if applicable)
	obtained outside of the health care system)."	EHR data can be subject to similar issues as claims, i.e., wrongly classified, incorrectly specified, or missing. Also, consider defining what constitutes a health care system. There are differing interpretations.[27]	
Line 161	"… how the data can be validated for a particular research activity."	PhRMA notes that not all data needs to be validated. Because the term '*validated*' has specific downstream implications (e.g., that additional validation studies need to be done on the specific study element), PhRMA recommends alternative wording.	PhRMA offers the suggested language: "and how the data can be ~~validated~~ **suitable** for a particular research activity."
Line 166-169	"Patients in different types of commercial or government health care payment programs can differ in a range of characteristics, such as age,	PhRMA suggests use of the term "*covariates*" rather than "*confounders*." Covariates include both confounders and	"Patients in different types of commercial or government health care payment programs can differ in a range of characteristics, such as age, socioeconomic status, health

[27] See Agency for Healthcare Research and Quality, Defining Health Systems. Available at https://www.ahrq.gov/chsp/chsp-reports/resources-for-understanding-health-systems/defining-health-systems.html.

Line No.	Existing Text	Comment	Proposed change (if applicable)
	socioeconomic status, health conditions, risk factors, and other potential ***confounders***."	effect modifiers as defined in the glossary.	conditions, risk factors, and other potential ~~confounders~~ ***covariates***."
Line 169-170	"Various factors in health care systems and insurance programs, such as medication tiering (e.g., first-line, second line), …"	PhRMA notes that "*medication tiering (e.g., first-line, second-line)*" is confusing. This reference appears to be to co-payments but can be confused with the "line" of therapy (e.g., first-line, second-line, etc.).	Edit to avoid confusion.
Line 201	"Enrollment and Comprehensive Capture of Care"	PhRMA suggests that FDA distinguish between closed claims sources and open claims sources in the guidance.	
Line 203	"Continuity of coverage (enrollment and disenrollment) should be addressed when using EHR and medical claims data sources…"	Rather than use the term "*Continuity of Coverage*" consider the use of "*longitudinal data*" as this term may be more precise and better describe what data are available for patients for a study.	Consider revising, "*Continuity of coverage*" to "*Continuity of coverage in the RWD*." Alternatively, consider if the term "longitudinal data" might be more precise for a study population than "*continuity*" or "*continuity of coverage*." Certainly, "*continuity of coverage*" may be useful terminology to describe the claims data source for insurance coverage. Whereas for studies, we consider "*longitudinally*" or "*longitudinal data*" terminology to describe what data are available for the patients. Lastly, consider using the term "*continuity of observable period*" to indicate the period

Line No.	Existing Text	Comment	Proposed change (if applicable)
			that patients remain in the healthcare system.
Lines 217 - 219	”A second example is a study where an outcome is dependent on a specific frequency of laboratory tests, and clinicians do not typically order those tests at such a frequency.”	The example provided in the documentation does not reflect the issue of comprehensiveness of the data sources in capturing aspects of care and outcomes that are relevant to the study question. In this case, the data source may comprehensively capture the outcome, but the outcome may still rarely show up as it is not a norm for clinicians to order those tests indicative of the outcome.	PhRMA recommends that the Agency include a more appropriate example of an outcome that is not captured in the medical claims data source such as laboratory results.
Lines 356-366	“Computable Phenotypes Standardized computable phenotypes can facilitate identification of similar patient populations and enable efficient selection of populations for large-scale clinical studies across multiple health care systems. A computable phenotype definition should include metadata and supporting information about the definition, its intended use, the clinical rationale or research	PhRMA notes that the term “*computable phenotype*” is not compatible with other terminology in the context of this guidance. A computable phenotype is essentially an operational definition for the target patient population (i.e., a code-based electronic algorithm using structured data elements). PhRMA recommends that FDA publish or acknowledge a	PhRMA recommends using the phrase “operational definition for the target patient population” instead of “computable phenotype” to maintain consistency with later sections of the guidance that address operational definitions for exposures, outcomes, and covariates.

Line No.	Existing Text	Comment	Proposed change (if applicable)
	justification for the definition, and data assessing validation in various health care settings…"	standardized list of computable phenotypes and code lists that will form a library of approved lists within particular conditions of interest.	
Lines 386-393	"All of these methods are computer-assisted to various levels but currently require a significant amount of human-aided curation and decision-making, injecting an additional level of data variability and quality considerations into the final study-specific dataset. If the protocol proposes to use AI or other derivation methods, the protocol should specify the assumptions and parameters of the computer algorithms used, the data source from which the information was used to build the algorithm, whether the algorithm was supervised (i.e., using input and review by experts) or unsupervised, and the metrics associated with validation of the methods. Relevant impacts on data quality	PhRMA requests that FDA clarify the intent of this section and provide more specificity regarding the use of the terms 'supervised' and 'unsupervised.' The terms 'supervised' and 'unsupervised' have specific meanings with the AI/ML field where supervised machine learning is a subcategory of ML and AI defined by its use of labeled datasets to train algorithms to classify data or predict outcomes accurately. Unsupervised learning trains algorithms using data points that are neither classified nor labeled. Clarity with these terms will help sponsors understand the intent of the recommendations and whether FDA is referring to the algorithm development or	FDA should define these terms in the glossary.

Line No.	Existing Text	Comment	Proposed change (if applicable)
	should be documented in the protocol and analysis plan."	human oversight of an AI-assisted curation process.	
Line 395; Section III. C.	"Missing Data: General Considerations"	PhRMA notes that "*missing data*" is an often-misunderstood issue and using it as a title for the section may perpetuate confusion. When there are missing data, it is always important to understand what the implications are for the study in regard to the missing data. Although FDA may address missing data on a case-by-case basis, this is such a critical issue for sponsors that FDA should address it directly in the guidance.	PhRMA recommends changing the title of this section to "*Unavailable Data*" or "*Gaps in Data*."
Line 441-453	"Because operational definitions are usually imperfect and cannot accurately classify the variable of interest for every subject, a resulting misclassification can lead to false positives and false negatives…"	PhRMA requests that FDA expand the current discussion to include more general comments on how measurement error and the performance of operational definitions can be assessed for variables that are not binary ,such as time-to-event and continuous variables. The discussion of the consequences of measurement	

Line No.	Existing Text	Comment	Proposed change (if applicable)
		error focuses on misclassification, and thus seems to implicitly assume that the variables of interest will be binary. Outcome variables may be continuous, count, or time-to-event, etc. General discussion of how the impact of measurement error can be assessed would therefore be helpful.	
Line 444	"Although complete verification of a variable of interest minimizes misclassification and maximizes study internal validity, understanding the implications of potential misclassification for study internal validity and study inference is the key step in determining what variables of interest might require validation and to what extent."	FDA seems to be interchangeably using "complete verification" and "validation" in this draft guidance, which is confusing.	PhRMA recommends that FDA use "validation" consistently throughout the draft guidance.
Line 445	"Although complete verification of a variable of interest minimizes misclassification and maximizes study internal validity, understanding the implications of potential misclassification for study internal validity and study inference is the key step in	The draft guidance states that complete verification of a variable of interest minimizes miscalculation and maximizes study internal validity.	

Line No.	Existing Text	Comment	Proposed change (if applicable)
	determining what variables of interest might require validation and to what extent."	It is not clear what is meant by "study internal validity" in the context of variable verification.	
Line 458	"(2) differential versus non-differential misclassification (e.g., differential misclassification of outcome by exposure); …"	The word "*exposure*" may mean much more than "*treatment groups*," e.g., dose, frequency, or route, here "*exposure*" may be better replaced by "treatment group." PhRMA recommends the following change "... differential misclassification of outcome by ~~exposure~~ treatment group …"	"(2) differential ~~versus non-differential misclassification (e.g., differential misclassification of outcome by exposure)~~ **misclassification of outcome by treatment group**…"
Line 500, Section V.	"Study Design Elements"	The title of this section, "*Study Design Elements*", may be confusing to researchers about its intent, when considering forthcoming guidances that will focus on study design choices, which this draft guidance does not discuss. Indeed, this section appears to be primarily focused on study variable definitions and how those are defined/validated in the RWD source.	PhRMA suggests an alternative title/nomenclature, such as "*Defining Study Variables*" or something similar.
Line 570	"The product of interest is referred to as the *treatment*, and	The draft guidance states that a possible comparator is placebo. It	"The product of interest is referred to as the *treatment*, and may be compared to no

Line No.	Existing Text	Comment	Proposed change (if applicable)
	may be compared to no treatment, a placebo, standard of care, another treatment, or a combination of the above."	is unclear how placebo as a treatment group would be assigned in an EHR or claims database analysis. PhRMA suggests deleting "*placebo*" or clarifying how this would be defined in a database study.	treatment, ~~a placebo,~~ standard of care, another treatment, or a combination of the above.
Lines 597 - 599	"When relying on coded data, the operational exposure definitions should be based on the coding system of the selected data source and reflect an understanding of the prescription, delivery, and reimbursement characteristics of the drug (if applicable) in that data source."	Some coding systems may change the code of a particular procedure/drug over time, or a specific code may be general and encapsulate many more specific procedures/drugs, with a shifting distribution over time. PhRMA recommends that FDA add "over time" at the end of the sentence (Lines 597-599).	"When relying on coded data, the operational exposure definitions should be based on the coding system of the selected data source and reflect an understanding of the prescription, delivery, and reimbursement characteristics of the drug (if applicable) in that data source **over time**."
Lines 599 - 602	"For example, in the United States, the operational definition should include the appropriate pharmacy codes (NDC or Healthcare Common Procedure Coding System) to capture the use of the drug in various settings."	Sometimes the Healthcare Common Procedure Coding System (HCPCS) use non-J codes.	FDA may want to consider this application and modify the sentence in Lines 600 - 601 to read: "(NDC or Healthcare Common Procedure Coding System)" to "(NDC or Healthcare Common Procedure Coding System or other applicable codes)."

Line No.	Existing Text	Comment	Proposed change (if applicable)
Lines 943 - 946	"Regarding outcome validation, sponsors should justify the proposed validation approach, such as validating the outcome variable for all potential cases or non-cases, versus assessing the performance of the proposed operational definition; if the latter will be done, justify what performance measures will be assessed."	PhRMA recommends that the FDA include an Appendix with specific examples of validation approaches, thresholds of sensitivity/specificity, etc.	
Lines 988-1003	"Information on potential confounders is collected in a nonrandomized study to support appropriate efforts to balance treatment and control groups in the analysis. […]"	PhRMA notes that the description of exposure should also comprise the description of the comparator time period in the case of a comparison with "*no treatment*" or with placebo.	PhRMA requests the definition of exposure be expanded to periods with no treatment, treatment interruptions, previous treatments, etc.
Lines 1001 - 1003	"FDA recommends considering potential linkages with other data sources or additional data collection to expand the capture of important confounders that are unmeasured or imperfectly measured in the original data source."	Linkages with other data sources may not always be possible.	PhRMA recommends additional language after the sentence ending on Line 1003, to read: "**If linkages with other data sources are not possible, the sponsor should discuss with the specific review division the appropriateness of using proxy variables to the specific unmeasured confounders in their study**."
Lines 1095, 1097, 1133	"Core data elements"	The draft guidance lacks clarity about the definition of "*core data elements*."	PhRMA requests "core data elements" be defined and differentiated from "key data element."

Line No.	Existing Text	Comment	Proposed change (if applicable)
Lines 1154-1187, Data Transformation	"Data transformation"	PhRMA requests that FDA add a reconciliation step between source data and CDM transferred data in the draft guidance. PhRMA believes this will decrease the chance of any missed matches or lag time.	
Line 1251	"VII. Glossary"	The guidance helpfully includes a glossary section but it does not include all the key terms so it would be beneficial if it was more comprehensive. PhRMA suggests that FDA include all key terms in the glossary for completeness. It is useful that the terms in italics are included in many instances, however, we provide some line-by-line specific terms as additional definition examples that FDA should include in the Glossary and offer some revisions to improve the definitions in the Glossary.	Additional examples to add: • Line 566 on page 14, consider including a definition for "Exposure" and defining what is a "discontinuation" in the Glossary. • For Line 1271 and Line 1350 from the Glossary, FDA introduce the definitions "*Conceptual definition*" at Line 760 and "*Operational definition*" at Lines 799 to 816 in this draft guidance. FDA should explain why these definitions are being used and offer practical examples to help explain the distinctions. It would be helpful if FDA expanded on why these are being used and use examples to better explain these terms in the Glossary as well as include references.
Line 1343	"Missing Data: Data that would have been used in the study	To the extent that the term "missing data" is used in the guidance, the current definition	PhRMA requests the definition of missing data be aligned with that given in the ICH E9(R1) Addendum.

Line No.	Existing Text	Comment	Proposed change (if applicable)
	analysis but were not observed, collected, or accessible."	of "*missing data*" given in the glossary is inconsistent with the definition given in ICH E9 Addendum.[28]	

[28] FDA/International Council for Harmonisation, Guidance for Industry - E9(R1) Statistical Principles for Clinical Trials: Addendum: Estimands and Sensitivity Analysis in Clinical Trials. Available at https://www.fda.gov/media/148473/download.