REAL-WORLD EVIDENCE ALLIANCE

info@rwealliance.org
rwealliance.org

April 7, 2025

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

Re: Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry; Availability; Comment Request (Docket No. FDA-2024-D-4689)

To the Food and Drug Administration:

The RWE Alliance appreciates the opportunity to comment on the draft guidance titled "Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products" ("Draft Guidance").[1] We are a coalition of real-world data ("RWD") and analytics organizations with a common interest in harnessing the power of real-world evidence ("RWE") to inform regulatory decision making to improve the lives of patients. Our members have deep knowledge and experience working with healthcare data across disease areas and patient populations, and we aim to bring these collective insights to bear in support of RWE policies.[2]

The RWE Alliance envisions a future in which data from electronic health records, administrative claims and billing records, product and disease registries, personal devices, wearables, and health applications will be used to generate evidence to support regulatory decision making related to medical product safety and effectiveness. To achieve these goals, the RWE Alliance advocates for policies that will (1) advance FDA's RWE Framework, (2) encourage the use of RWE to better understand treatment effects in underrepresented populations, (3) enhance opportunities for RWE

[1] 90 Fed. Reg. 1157 (Jan. 7, 2025).

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

organizations to consult with FDA, (4) increase communication on the generation and use of RWE, and (5) recognize the unique aspects of and opportunities for RWD/E.[3]

We appreciate FDA's effort to provide guidance on the use of artificial intelligence ("AI") to support regulatory decision making in a wide variety of contexts, including by processing and analyzing large sets of RWD to develop clinical trial endpoints or assess outcomes. The use of RWD in the development and deployment of AI technologies in the healthcare and life sciences sectors has the potential to yield enormous benefits. AI tools are rapidly expanding the types of RWD that can be collected and the methods for that collection, as well as the ability to analyze data and evaluate data quality. The use of high-quality, reliable, and representative RWD to train, improve, and evaluate AI systems used in medical product development meaningfully advances medical innovation for patients. Moreover, AI developed using RWD can accelerate the discovery, research, development, and delivery of advanced and personalized treatments and cures.

Section I of this letter provides general comments on the Draft Guidance, and Section II provides comments on specific portions of the Draft Guidance.

I. General Comments on the Draft Guidance

We offer general suggestions as FDA continues to build experience with AI and the research community develops new best standards and practices for AI's use in regulatory decision making.

First, we commend FDA for describing the potential for bias depending on the quality, size, and representativeness of datasets for training AI models, and for describing how to identify and mitigate potential sources of bias. Unlocking the potential of AI depends on the quality of the data used to build and evaluate the technologies. RWD is uniquely positioned to address bias. By providing information about the health status and outcomes of individuals in diverse care settings, RWD can be used to avoid problems associated with over- or under-representation, thus allowing for the development of generalizable AI models. We recommend that FDA continue placing emphasis on the importance of diverse, high-quality datasets and the necessity of mechanisms for detecting and mitigating AI bias, as such biases could impact clinical outcomes significantly.

Second, we agree with FDA that data used to develop AI models should be fit for use, meaning the data should be both relevant and reliable. While we commend FDA for referencing existing RWE-related guidance to define these terms and believe that it is important for the Agency to apply consistent expectations to its review of RWE, the Agency should clarify how it would evaluate reliability in the context of data used to train

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

AI models. The concept of reliability from FDA's RWE-related guidance—defined by the three prongs of accuracy, completeness, and traceability[4]—applies well to clinical studies because it evaluates whether RWD used to generate RWE is fit for use to support regulatory decisions of safety and effectiveness. In the context of evaluating the reliability of training data, these three prongs should be assessed under a risk-based approach to account for the wide range of use cases for AI models. FDA also should account for other approaches to assess reliability in the context of training data. For instance, sponsors and FDA may be able to overcome uncertainty regarding the reliability of training data in some circumstances by testing the performance of the AI model. This could take the form of pilot implementation data, observational study outcomes, or simulated-use studies demonstrating performance in settings that closely resemble the intended use environment. We suggest that FDA expand upon the Draft Guidance's point on reliability in the context of RWD used to train AI, including by clarifying ways to demonstrate reliability under the RWE-related guidance framework[5] and the relationship between an evaluation of data reliability and the relevant components of the credibility assessment framework. We also urge FDA to provide examples of reliability assessments for a range of low- to high-risk AI models for specific regulatory questions.

Third, the Draft Guidance does not account fully for emerging AI technologies. We suggest FDA expand its sections on credibility assessments and life cycle management to include examples for self-learning models, foundation models, and generative AI. These examples should illustrate the steps involved with creating a risk-based credibility assessment and address any unique considerations due to the AI model's design. For instance, self-learning models have many promising applications because they can learn and adapt continuously from new data, even after deployment, without needing retraining from scratch. Ensuring the credibility of the output of these models can raise unique considerations because their behavior changes over time as they learn from more data and they might not always follow a consistent or predictable pattern; as a result, specific types of controls or assessments may be appropriate.

Fourth, FDA should more explicitly consider transparency in the context of AI. FDA's current framework regarding the transparent documentation of training datasets, model validation, and interpretability is vague. We suggest that FDA offer recommendations

---

[4] FDA, Guidance for Industry: Real-World Data: Assessing Electronic Health Records and Medical Claims Data to Support Regulatory Decision-Making for Drug and Biological Products 3 (July 2024), https://www.fda.gov/media/152503/download.

[5] The RWE Alliance has commented extensively on the reliability framework in FDA's RWE-related guidances. *See, e.g.*, RWE Alliance, Comment on Public Workshop on Optimizing the Use of Real-World Evidence in Regulatory Decision-Making for Drugs and Biological Products—Looking Forward (Jan. 13, 2025), https://rwealliance.org/wp-content/uploads/2025/03/RWE-Alliance-Response-to-Request-for-Comments-on-RWE-Public-Workshop-1-13-25.pdf.

for how researchers should document and assess transparency, defined in terms of model behavior, of a complex model to satisfy any applicable regulatory requirements.

Fifth, we recommend that FDA provide specific details on the processes by which companies can engage with FDA during the development and training phases of AI models intended to support regulatory decision making, including on questions related to the selection of the data used to develop an AI model. We appreciate FDA directing the Draft Guidance to "Industry and Other Interested Parties" and welcome FDA continuing to engage with stakeholders in the RWE ecosystem when developing new policies on AI and RWD.

II. Comments on Specific Sections of the Draft Guidance

We provide our comments on specific sections of the Draft Guidance below.

- In lines 152-207, the Draft Guidance should provide recommendations on how to identify the regulatory decision and applicable regulatory standards associated with the question of interest and context of use. How the RWD is used will impact how sponsors shape both the question of interest and context of use and ultimately how the AI model risk is assessed.

- We agree with the Draft Guidance that data used to train and tune an AI model during the development phase should be fit for use and that testing data should not only be sufficiently fit for purpose but also independent of the development data. Beyond the data used for development and testing, however, the credibility assessment also should consider factors related to real-world deployment of an AI model used to generate data for regulatory decision making. We recommend FDA add considerations regarding factors that should be incorporated routinely into the evaluation of AI model uses, especially for those deemed to be higher risk as the decision consequence or model influence increases. Credibility assessments should consider real-world workflow considerations, human factors, and other aspects of usability testing in settings that closely resemble the intended use environment.

- In lines 323 to 404, the Draft Guidance's assumptions for plan development envision traditional, non-generative AI solutions where the AI model is trained upon the given dataset using a defined evaluation metric. However, these assumptions may not fit well for newer generative AI models that use Retrieval Augmented Generation ("RAG") techniques and agentic approaches. RAG tools can enhance the accuracy and reliability of generative AI models by analyzing information outside of the training data sets used by large language models. To evaluate the robustness of an AI system that incorporates a RAG or agentic framework, it is useful to compare its performance to a base model without these components across a variety of scenarios, including challenging queries, ambiguous information, and diverse data sources. Such tests would use metrics like retrieval accuracy, response relevance, faithfulness to the retrieved context,

and consistency in response generation. We ask FDA to discuss these techniques in the final guidance and address other relevant AI techniques and methodologies as the field evolves.

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

Best regards,

The RWE Alliance