April 7, 2025

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

RE: Docket No. FDA-2024-D-4689; Considerations for the Use of Artificial Intelligence To Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry and Other Interested Parties

To Whom It May Concern:

On behalf of LUNGevity Foundation, the nation's preeminent lung cancer nonprofit that funds research, provides education and support, and builds communities for the more than 230,000 Americans diagnosed with lung cancer each year[i] and over 600,000 Americans living with the disease[ii], we appreciate the opportunity to submit these comments to the U.S. Food and Drug Administration (FDA) regarding the Draft Guidance "**Considerations for the Use of Artificial Intelligence to Support Regulatory Decision-Making for Drug and Biological Products.**"

LUNGevity is committed to and actively engages in efforts to improve the efficiency with which clinical research is conducted. Leveraging artificial intelligence (AI) tools throughout the development and testing of drug and biologic products can optimize the development process and accelerate the delivery of safe and effective therapies to patients with lung cancer. It is imperative that AI tools used in drug development (and other applications impacting patient care) undergo adequate validation relevant to their intended use to ensure outputs are reliable and applicable to all patients. We thank the FDA for issuing this draft guidance and we offer the following recommendations as the Agency finalizes the document.

Early Consultation with FDA on AI Model Development

LUNGevity appreciates the inclusion of Section IV.C which guides sponsors on whom to contact at the FDA on expected credibility assessment activities based on the intended use of an AI model. Guidance on early engagement with the Agency will allow sponsors to effectively plan activities and appropriate assessments, facilitating more rapid implementation and validation of their AI models in the generation of data for regulatory decision-making. However, we encourage the FDA to also provide guidance on how best to engage with the Agency even earlier, during AI model development, rather than only after

the model has been built and the appropriate credibility assessment activities are being determined. This early discussion with the Agency could potentially prevent the development of models with inadequate training or validation datasets.

In addition to opportunities for consultation with the FDA, sponsors could benefit from greater clarity in the final guidance on considerations for AI model development, particularly in ensuring the training data for a proposed AI model are fit for use. For example, the draft guidance encourages the use of data used to develop AI models that includes "sufficient numbers of representative participants." We recommend that the Agency elaborate on what it would base its determination of sufficient representativeness, in addition to advising sponsors on how they could confer with the Agency on this and other concerns during AI model development in the final guidance.

Achieving Methodological Transparency in AI Model Development

The draft guidance highlights the importance of methodological transparency and encourages the inclusion of details surrounding the methods and processes used to develop an AI model in regulatory submissions. We recommend providing further specificity on the types of information surrounding the methods and processes used in the AI model's development that the Agency would expect. The Agency may consider developing an optional template to aid in gathering the relevant information akin to the proposed HARPER+ protocol template[iii] drafted by the Centers for Medicare and Medicaid Services (CMS) for use in Coverage with Evidence Development (CED) protocol submissions using real-world data (RWD). Similar to the HARPER+ template, which solicits information on data management procedures, quality control measures, and other aspects of RWD studies, an optional template requesting pertinent information on AI model development (e.g., training data set characteristics, etc.) could provide clarity to sponsors on what information to include in the regulatory submission.

Furthermore, while clinical trial sponsors may test the performance of an AI model for its intended use in a trial, the model may be developed by a contracted external vendor from whom methodological details may be difficult to obtain. Having a set template clearly outlining the pertinent information needed on an AI model's development could facilitate the collection of all necessary information from all parties needed for the regulatory submission. This mirrors the HARPER+ template which CMS notes may require input from RWD vendors.

Defining the Context of Use of an AI Model

Per the draft guidance, the context of use (COU) of an AI model encompasses its specific role in addressing the question of interest and its contribution in doing so relative to other sources of information (if any) within a distinct development program. However, there may be instances where an AI model may serve the same role in different contexts, for example, across different drug development programs. The draft guidance does not provide considerations for when it may or may not be appropriate to extrapolate the results of one credibility assessment for a given COU to other similar COUs. Additionally, it is unclear if the COU of an AI model could be defined in a way that could apply its use to multiple development programs. We request additional clarity on how broadly a COU may be defined, as well as the credibility assessment requirements in cases where a COU might apply across development programs (e.g., expectations regarding demonstration of applicability of training and validation data sets across different programs).

Assessing Risk of an AI Model

The draft guidance defines the risk of an AI model based on its relative influence and decision consequence, providing one example of a risk assessment of an AI model used in clinical development. The example provided is defined as high risk, based on the high probability that an incorrect decision would lead to an adverse outcome, and that no other sources of information are used to address the question of interest. Additional examples across the spectrum of risk (e.g., not as clearly high or low) in both relative influence and decision consequence for AI model risk assessment in both clinical development and commercial manufacturing, or providing additional clarification or factors for defining risk across model influence and decision consequence, would be helpful for sponsors to adequately assess their models' risk.

AI Model Performance Evaluation

As part of the credibility assessment of an AI model used to support regulatory decision-making, sponsors are asked to describe the reference method used to test the model's performance, as well as a summary of the reference method's performance. As the reliability of the reference method is critical to accurately evaluating the AI model's performance, we recommend the draft guidance provide further details on the information the Agency expects to adequately ensure the reference method is appropriate. For example, if the reference method involves human decision-making, interobserver variability can impact the performance assessment and, along with representativeness of the observers, should be considered in the evaluation of the reference method.

The draft guidance also recommends the use of evaluation methods that consider the performance of the human-AI team in cases where the COU involves a "human in the loop."

As noted above, humans will also introduce variability, and therefore further clarification on the Agency's expectations for evaluation of the human-AI team is needed, including ensuring assessments include the spectrum of humans representative of those in the context of use to adequately evaluate the impact of variability on performance associated with a human in the loop. Further elaboration on the evaluation methods the Agency may expect would also be helpful. For instance, the final guidance could comment on whether standalone performance of both the model and the human component should be assessed as part of the evaluation, where applicable.

Further, while the draft guidance outlines the performance metrics that should be used to evaluate their AI models (e.g., sensitivity, specificity), it remains unclear whether the Agency will expect a certain threshold of minimum performance based on those metrics to be met, and how sponsors will be expected to comply. The FDA may, for instance, set expectations for model performance at the stage of the risk-based credibility assessment wherein the sponsor develops their plan to establish AI model credibility (step 4), and expect those performance thresholds to be met upon execution of that plan (step 5).

Lastly, as noted in section IV.B, it is critical to ensure that the model remains fit for use over time. The draft guidance notes that some steps of the credibility assessment may need to be re-executed. Further guidance is needed to determine when this re-execution is necessary, how the results should be reported, and what action the Agency may take if the performance of the model has changed significantly such that the model may no longer be fit for use. Appropriate continued assessment is necessary to ensure the model is performing as intended for the context of use to ensure patient safety.

LUNGevity appreciates the opportunity to comment on this important guidance. Outlining regulatory expectations around the use of AI models to support regulatory decision-making will help this technology reach its potential to accelerate drug development by ensuring such tools are adequately fit for use. In addition to the clarifications proposed here, we also encourage the FDA to ensure that the definitions, recommendations, and considerations in the finalized guidance are harmonized with those of other regulatory authorities to optimize applicability in multi-regional clinical trials. Please feel free to reach out to me at bmckelvey@lungevity.org with any questions.

Sincerely,

Brittany McKelvey

Brittany Avin McKelvey
Senior Director, Regulatory Policy

On Behalf of LUNGevity Foundation

---

[i] Howlader N, Noone AM, Krapcho M, et al. (eds). SEER Cancer Statistics Review, 1975-2018, National Cancer Institute. Bethesda, MD, https://seer.cancer.gov/csr/1975_2018/, based on November 2020 SEER data submission, posted to the SEER web site, April 2021.

[ii] Centers for Disease Control and Prevention. United States Cancer Statistics. Available at https://gis.cdc.gov/Cancer/USCS/#/Prevalence/

[iii] Centers for Medicare & Medicaid Services. Modified HARmonized Protocol Template to Enhance Reproducibility – HARPER+; Proposed Guidance Document: Study Protocols That Use Real-World Data. https://www.cms.gov/files/document/appendix-rwd-proposed-guidancepdf.pdf