AETION

January 24, 2022

Dockets Management Staff (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Rm. 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Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry"

To the Food and Drug Administration:

Thank you for publishing the draft guidance document entitled "Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products; Draft Guidance for Industry" (FDA-2020-D-2307) (the "Draft Guidance"). Aetion appreciates the Agency's commitment to advancing the use of real-world data (RWD) and real-world evidence (RWE) in regulatory decision-making. Integrating RWE into regulatory decision-making will facilitate more efficient drug development, enhance understanding of product safety and effectiveness, and help support the shift to value-based care. And the Agency's work to provide robust, actionable guidance will help to better ensure that regulatory submissions containing RWD and RWE meet FDA's appropriately high expectations.

Aetion welcomes and appreciates the opportunity to comment on the Draft Guidance. As a health technology company that specializes in the generation of RWE capable of supporting regulatory decision-making, we confront the matters addressed in the guidance on a daily basis. Our flagship product, the Aetion Evidence Platform® (AEP), was originally created by two academia-based pharmacoepidemiologists to address a widely recognized need: scalable, transparent, and scientifically rigorous analyses of RWD to identify the safety, effectiveness, and value of clinical interventions.

Aetion shares FDA's commitment to fostering high-quality analyses of RWD that can provide evidence for regulatory decisions across the product life cycle, including regulatory review of safety and effectiveness in the context of product approval decisions. The Draft Guidance is a critical next step to advancing FDA's RWE Framework and the use of high-quality RWE in regulatory decision-making. Greater alignment on principles and best practices, and greater public visibility regarding FDA's expectations, will enable sponsors and other researchers to approach the analysis of RWD with greater confidence that such efforts will lead to the generation of valuable evidence. Further, integrating transparent, auditable, reproducible, and scientifically valid RWE into regulatory decision-making will facilitate more efficient drug development and enhance understanding of product safety and effectiveness.

Overall Comments

For the reasons described above, we very much appreciate the Draft Guidance's elaboration of key principles that the Agency has identified as necessary components of RWE studies capable of supporting FDA decision-making. We strongly agree that RWE studies should include elements to ensure that the RWD being used is reliable, relevant, and fit-for-purpose, and that the study is designed to be transparent and reproducible.

We also agree with the Draft Guidance's general approach of encouraging the pre-definition of essential elements in protocols and where relevant, statistical analysis plans (SAPs). Conducting studies using these principles helps to address concerns around potential bias and confounding, and helps to ensure that RWE analyses will produce interpretable and clinically meaningful results.

We particularly appreciate the level of nuance with which the Agency has approached important considerations in the guidance. For example, in Section IV.C, which provides general considerations about missing data, the Draft Guidance discusses the important distinction between implicit and explicit missing data. This consideration is not consistently addressed by researchers and we appreciate FDA's attention to the topic. Likewise, on lines 934-937, the Draft Guidance notes that two dependent misclassifications can create a bias away from the null — a nuanced consideration that we are pleased to see addressed.

Implementing the recommendations in the guidance will require substantial care and attention to detail, but we believe these recommendations are generally achievable with the use of currently available best-practices, scientific expertise, and technology. For example, in lines 742-743, FDA states "Whether and to what degree a data source captures the outcome of interest should be assessed before study initiation and be independent of the exposure of interest." A software platform like AEP can help achieve such transparency through features such as automated audit trails. A software platform can also facilitate checks to confirm that the choice of a reference standard, validation approach, methods, processes, and sampling strategy align with the pre-specified protocol. Additionally, a platform can ensure adequate data are available to FDA and other reviewers to assess the comparability of the exposed and comparator populations, and to check data for completeness and document any transformation made during the process of ingesting data into the platform.

While a software platform can address many of the Draft Guidance's recommendations, there are some aspects that are beyond the scope of a sponsor and an analytics platform to address without information from the data source. We agree that it is important to document processes that are upstream to selection of a data source by a sponsor and ingestion into a software platform — including, for example, transformations to the source data, provenance, and accuracy and completeness of the data attributes, and the manual curation process to create a clinical data repository. However, without information from the data provider or data source, it may not be feasible for a sponsor or analytics provider to provide or reference documentation that could be included in a submission (or referenced during the review process). To that end, the extent to which a data provider makes such information available

may be a relevant consideration when a sponsor or analytics provider assesses whether a particular dataset is fit for use in a planned study. We therefore believe it would be helpful if FDA, when finalizing the Draft Guidance, could provide recommendations for how sponsors and analytics providers should evaluate the level of available information when deciding whether to use a commercially available dataset in a study.

Study-Specific Considerations

It is important to recognize that not every recommendation in the guidance will be feasible or appropriate in all studies, and we are encouraged that the Draft Guidance in many places explicitly acknowledges that recommended approaches may need to be adapted based on the context of the study (e.g., lines 463-469, 821-823, 888-890). We believe it is important to recognize that the best approach may vary depending on the context of the specific study, and we encourage the Agency to further emphasize and describe its willingness to consider a clearly rationalized, tradeoff-based approach to the overall validity of a given study. To that end, we recommend the Agency align the language in the Draft Guidance to the approach in the draft guidance "Real-World Data: Assessing Registries to Support Regulatory Decision-Making for Drug and Biological Products," in which FDA repeatedly clarifies its flexible approach based on study- and context-specific considerations.

In this vein, we believe additional demonstration projects or Advancing Regulatory Science research will further develop and refine pharmacoepidemiological approaches. The Agency's Advancing Real-World Evidence for Use in Regulatory Decision-Making program under the Prescription Drug User Fee Act VII Commitment Letter will provide strong insights into developing regulatory-grade RWE, but some of the considerations in this guidance offer additional areas for methodological development. For instance, in lines 609-611, FDA discusses extraction of unstructured data, a cutting edge topic in this field. We recommend demonstration projects to fully explore the use of this kind of data, determine standards, and evaluate the potential of unstructured data to support regulatory decision making. Additionally, in lines 692-694, the Agency states that it recommends documenting the methods used to calculate and validate medication switching. Switching is a topic of great debate and currently, there is no validated method for clearly identifying medication switches. Demonstration projects or additional research could lead to a stronger understanding of this aspect of pharmacoepidemiology.

Data Linkage and Synthesis

We appreciate the Draft Guidance's focus on data linkage and synthesis (lines 248-288), as linking is an important tool that can widen the aperture of information gleaned from RWE. We believe that strong and explicit guidance from the Agency on this topic will help to drive methodologically sound approaches. To that end, we recommend that the Agency clarify this discussion in certain respects in order to maximize its utility to study sponsors and to avoid inadvertently discouraging study designs that may be most appropriate in context.

In particular, while we appreciate the discussion in the Draft Guidance highlighting the distinction between deterministic and probabilistic approaches to data linkage, we believe that differences in the Draft Guidance's treatment of the two approaches could lead to unintended consequences. As drafted, the Draft Guidance recommends that a protocol using the probabilistic approach — but not one using the deterministic approach — should include testing the impact of the degree of match and robustness of findings as part of the analysis plan. That this differential burden could inadvertently discourage the use of probabilistic approaches in appropriate circumstances, and may be perceived as a broad Agency preference for deterministic linkages. We believe the benefits of linkage — regardless of whether a deterministic or probabilistic method is used — outweigh the absence of linkage. In our members' experience, both deterministic and probabilistic linking, when designed and executed correctly, can be effective depending on the research question. What matters most is how the underlying methodology is executed, not the choice of methodology itself. We suggest the Agency clarify this discussion to reinforce the importance of well-executed methodological approaches, regardless of the specific methodology used.

Separately, with respect to heterogeneity, at lines 269-272, the Draft Guidance states:

"For studies that require combining data from multiple data sources or study sites, FDA recommends demonstrating whether and how data from different sources can be obtained and integrated with acceptable quality, given the potential for heterogeneity in population characteristics, clinical practices, and coding across data sources."

This statement could be misconstrued as suggesting that heterogeneity in population characteristics and clinical practice should be avoided. To the contrary, we believe that such heterogeneity has considerable benefits. For example, it can improve the generalizability of a study across health care systems and health care plans by increasing representation of prescribing or care practices in multiple systems or geographic areas, or formulary practices. Additionally, heterogeneity in population characteristics can provide additional insight into causal effects within traditionally underrepresented groups.

As the guidance points out, heterogeneity should be accounted for in study design. For example, when it comes to heterogeneity in population characteristics, researchers can stratify the study population by age, race, or other demographic indicators. Likewise, heterogeneity in coding practices can be addressed by designing the protocol or SAP to capture multiple related codes that may be used to code the same event (e.g., myocardial infarction) across different providers or systems, instead of focusing on a single code.

We recommend revising the statement at lines 269-272 to acknowledge the benefits of heterogeneity and recommended approaches for accounting for it in study design.

Distributed Data Networks and Common Data Models

We appreciate the Draft Guidance's focus on distributed data networks and Common Data Models (CDMs) (lines 290-354). CDMs can be a useful tool for RWE analysis, but we believe the choice of CDM can have significant implications for the validity of an RWE study. Accordingly, we believe it would be helpful for the Agency to provide additional guidance regarding the differences between types of CDMs and their implications for study validity.

In particular, it is important to differentiate between organizing CDMs, which organize data to facilitate queries across datasets while preserving the underlying source data, and CDMs that map the data onto a standardized set of parameters, thus modifying the source data (sometimes referred to as preconfigured rules systems). Whereas organizing CDMs are are fully adaptable to different types of research and are appropriate for use in regulatory decision making, mapping data onto preconfigured rules can lead to difficulties in study consistency and validity. To promote greater reliability and relevance in RWE studies, we recommend that the Agency recognize this distinction and its implications for regulatory decision making. In particular, it would be helpful for the Agency to acknowledge specifically that using a mapping CDM may lead to the loss of underlying data and thus may reduce the validity of the study.

Glossary and Key Terms

FDA defines "Study Period" as "The calendar time range of data used for the study (Wang et al. 2017)."

We largely agree with this definition, but suggest the Agency modify it to "The calendar time range of data used for the study (Wang et al. 2017), including any necessary lookback period." The definition as currently written may imply that study period does not include a lookback period, which is an important component of observational study design. Indeed, this point is consistent with lines 221-223: "FDA recommends specifying how all relevant populations, exposures, outcomes, and covariates will be captured during the study period, particularly in situations where data availability varies greatly over time."

Line Item Comments

Lines 97-99: *"For all studies using EHRs or medical claims data that will be submitted to FDA to support a regulatory decision, sponsors should submit protocols and statistical analysis plans before conducting the study."*

If a protocol contains sufficient detail on the design, analysis, and operational definitions of study variables (the latter of which may be put in an appendix to the protocol), a separate analysis plan is not necessary. We suggest clarifying in the text that the protocol and if needed, any other documentation (such as a statistical analysis plan or an implementation plan) should be submitted to FDA for review.

Lines 99-100: *"Sponsors seeking FDA input before conducting the study should request comments or a meeting to discuss the study with the relevant FDA review division."*

It would be helpful to indicate how sponsors should seek FDA input on protocols prior to implementation. Sponsors typically use Type B and C meeting requests. We suggest that FDA specify whether these meeting requests or any other paths (including new paths that may be made available under the next user fee agreement) are most appropriate for seeking FDA input.

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 obtained outside of the health care system)."*

We agree with the considerations for assessing an EHR dataset outlined in this section. We suggest FDA expand this section to discuss how longitudinality and observability of EHR data can impact study validity. Understanding the longitudinality of data is a crucial component of working with EHR data, as it may be limited (especially in the US).

Additionally, observability is a key concept that should be addressed in this section of the guidance. If a patient's observability status is unknown, then we do not know if the absence of a health care event (e.g., a stroke) means the event did not happen — or that it happened but was not recorded. Thus, it is important to know the time intervals over which a patient is observable so that appropriate analytical methods can be applied to deal with periods of unobservability. Claims datasets often provide excellent information on observability because they explicitly define periods of time when patients are enrolled in health insurance and are thus contributing relatively complete data on their health care encounters. EHR datasets, in contrast, generally do not indicate when patients can be seen at the provider network from which the EHR was derived—thus observability needs to be dealt with more carefully. We suggest that the limitation regarding observability when using EHR is noted in the guidance, along with the value of linking EHR data to claims data, which can be used to more precisely identify unobservable periods.

Lines 152-156: *"For prospective clinical studies proposing to use EHRs, it may be possible to modify the EHR system for the purpose of collecting additional patient data during routine care through an add-on module to the EHR system. However, given the limited ability to add modules to collect extensive additional information, EHR-based data collection may still not be comprehensive."*

This is cited as an example of a limitation arising from the fact that electronic healthcare data were not developed to support regulatory submissions to FDA. We agree that the ability to add fields /modules to an existing EHR system is rare; however, if it is possible, it likely

improves study validity. We suggest replacing this example with another example that is more common and likely to lead to bias, such as capture of a procedure occurrence but not the clinical finding from the procedure (e.g., pathology findings from a biopsy).

Lines 223-226: *"The data sources should contain adequate numbers of patients with adequate length of follow-up to ascertain outcomes of interest based on the biologically plausible time frame when the outcome, if associated with the exposure, might be expected to occur."*

We suggest this sentence be modified to read "The data sources should contain adequate numbers of patients with adequate length and continuity of follow-up to ascertain outcomes of interest based on the biologically plausible time frame when the outcome, if associated with the exposure, might be expected to occur." Length of follow-up alone may not capture all related outcomes. Researchers should also assess the continuity of follow-up to ensure all relevant events are captured.

Lines 251-255: *"If the study involves establishing new data linkages between internal data sources (e.g., mother-infant linkages) or external data sources (e.g., vital records, disease and product registries, biobank data), the protocol should describe each data source, the information that will be obtained, linkage methods, and the accuracy and completeness of data linkages over time."*

It is unclear what "linkages between internal data sources (e.g., mother-infant linkages) or external data sources (e.g., vital records, disease and product registries, biobank data)" refers to. Does "linkages between internal data sources" refer to linkages within a single data collection infrastructure? Does "linkages between… external data sources" mean linking one dataset (e.g, EHR) to another data source from a different data collection infrastructure (e.g., a registry)? Clarifying this statement will aid reader comprehension of the guidance.

Lines 255-259: *"If the study involves generating additional data (e.g., interviews, mail surveys, computerized or mobile-application questionnaires, measurements through digital health technologies), the protocol should describe the methods of data collection and the methods of integrating the collected data with the electronic health care data."*

We suggest the Agency modify this sentence to read "If the study involves generating additional data (e.g., interviews, mail surveys, computerized or mobile-application questionnaires, measurements through digital health technologies), the protocol should describe the methods of data collection and the methods of integrating the collected data with other sources of real-world data." We recommend broadening the last clause, as this does not necessarily only apply to integrating collected data with EHR data.

Lines 358-365: *"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 justification for the definition, and data assessing validation in various health care settings (Richesson et al. 2016). The computable phenotype definition, composed of data elements and phenotype algorithm, should be described in the protocol and study report and should also be available in a computer-processable format."

From the description of "computable phenotype" given here and the definition given on lines 1266-1268, it's not clear how this differs from a typical algorithm used to identify a condition or event of interest. We suggest providing additional clarification.

Lines 397-401: "There are two broad cases in which information may be absent from the data sources. The first case is when the information was intended to be collected (e.g., structured field present in the EHR), but is absent from the data sources. This is an example of traditional missing data. The second case is when the information was not intended to be collected in the EHR and medical claims data and is therefore absent."

We appreciate the FDA's distinction between the two types of missing information, which is important when undertaking a feasibility assessment for a particular research question. The second type of missingness is helpful to reduce, at the outset, the number of candidate data sources for further consideration. The first type of missingness can then be assessed among the retained candidate data sources. In our experience, these types can be confused and labels may improve communication between researchers and data owners. As a suggestion, the first case may be labeled as "patient-level missing data" and the second case as "data source-level missing data."

Lines 463-469: "Although complete verification of a study variable is considered the most rigorous approach, there are scenarios where verifying a variable for every subject might not be feasible (e.g., a very large study population, lack of reference standard data for all study subjects) and assessing the performance of the variable's operational definition might suffice. Based on the performance measures described in Table 1, sponsors should consider whether validating the variable to a greater extent (e.g., all positives classified by the operational definition) is necessary and discuss with the relevant review division."

We appreciate the acknowledgement that complete verification may not be feasible when the study population is very large, etc. We suggest clarifying the last sentence to provide additional options in the instance where the variable definition performance (assessed via the measures described in Table 1) is deemed unacceptable; these options may include updating the variable's definition and validating this updated definition against a reference standard in a new sample of provisional positives. In addition, we suggest removing the example in parentheses, which seems to describe complete verification as it refers to validation of "all positives" rather than a sample. Moreover, complete verification or even validation on a sample of provisional positives is often infeasible due to patient privacy. We suggest that the guidance acknowledge additional flexibility for situations in which there is a need for evidence, even with imperfections and other means to improve interpretability can

be undertaken, such as using analogous validated algorithms (e.g. validated in another data source of similar type), and conducting sensitivity analyses.

Lines 545-550: *"Key variables used to select the study population should be validated. For example, to assess the drug effect in patients with immune thrombocytopenic purpura, the disorder ascertained by operational definition International Classification of Diseases, Ninth Revision, Clinical Modification (ICD-9-CM) diagnosis code 287.31 should be validated based on the conceptual definition of the disorder, which includes signs and symptoms, levels of platelets, and exclusion of other possible causes of thrombocytopenia."*

It is unclear to which entities the recommendations in this section are directed. As drafted, it could be interpreted as suggesting sponsors should validate ICD codes themselves, which is typically infeasible. The use of these codes to capture the clinical outcome of interest involves complex processes to build measures, definitions, and algorithms (the measures). We recommend the Agency clarify this section to state that protocols, statistical analysis plans, and other related documents should specify the assumptions and parameters of the measures, the data source(s) used to build these measures, and any validation and metrics associated with the measures.

Lines 556-558: *"If such data are used, the protocol should describe the source of information and the methods health care providers use to generate the data (if known)."*

The Agency should consider omitting the clause "the methods health care providers use to generate the data (if known)." This information is rarely, if ever, known.

Lines 629-632: *"The data source should capture the relevant exposure duration (anticipated use of a product over time). Given that some medical products are designed as one-time exposures (e.g., vaccines), and other products may be intended for use over extended periods of time, the suitability of a data source will vary with the specific medical product under investigation."*

We agree that duration of an exposure is an important consideration when assessing RWD. However, a record of a drug being administered does not necessarily mean a patient is using the drug in accordance with the prescribing information captured in its label. We suggest FDA broaden this section to include not only product characteristics, but conditions of use and conditions of the underlying disease. All of these confounders may affect the duration of exposure variable. For instance, drugs given intravenously are likely being used as intended, but there may be less confidence in the actual use of inhaled products. Whether a disease is acute or chronic can also affect this measurement.

Lines 668-671: *"Other than for medications administered in hospital settings or infusion settings, electronic health care data capture prescriptions of drugs and the dispensing of drugs to patients, but generally do not capture actual patient drug exposure because this depends on patients obtaining and using the prescribed therapy."*

We agree that actual patient drug exposure is not typically captured in EHR data. Actual drug use may be reasonably inferred by regular refillings/dispensings aligned with days supply, and analytic rules about allowable gaps between refillings/dispensings can be implemented with sensitivity analyses to assess their impact on results. Additionally, we believe that there are some settings where EHR data can provide this information. Additionally, linking to an external data source, such as a patient registry, could provide this information. We ask that the Agency rephrase this section to capture this possibility.

Lines 790-797: *"The sensitivity and specificity of an operational definition are imperfect when there is outcome misclassification. Given that it is usually not possible for sensitivity and specificity to be perfect (i.e., 100%), outcome misclassification might result in both false positives and false negatives. FDA recommends considering the potential impact of outcome misclassification on study validity when developing or selecting an operational definition for the proposed study. For example, when studying infrequently occurring outcomes in a cohort study, given the low prevalence of the outcome event, it is important to achieve high specificity to minimize false positive cases and high sensitivity so that more true cases can be captured."*

We agree that sensitivity and specificity analyses may not be perfect in a given study and considering the potential impact of outcome misclassification is an important step when designing an RWE study. However, we believe this section could be expanded to capture additional pharmacoepidemiological nuances. For instance, when estimating a risk ratio, specificity is more important than sensitivity; the risk ratio will be valid even if the outcome measure has some false negatives from imperfect sensitivity and regardless of outcome prevalence, as long as there are no false positives (perfect specificity). Thus, focusing on high specificity in this situation will help ensure the resulting ratio is correct. We absolutely agree that these parameters are important to measure, but we suggest they be further described in light of the idea that certain data can be fit for purpose and inexact.

Lines 853-866: *"PPV is often assessed in validation studies. PPV is the proportion of potential cases identified by an operational definition that are true-positive cases. Therefore, PPV informs the degree to which false-positive cases are included among the identified cases. When the concern with false-negative cases is negligible (e.g., when the sensitivity is deemed sufficiently high so that the number of false-negative cases is minimal), a high PPV might be adequate to provide confidence in the validity of the outcome variable, whereas a moderate-to-low PPV might warrant complete verification of the outcome variable for all potential cases. When the extent of false-positive cases and the extent of false-negative cases are of concern, sponsors should consider assessing all performance measures needed for quantitative bias analysis to evaluate the impact of outcome misclassification on the measure of association or take a more rigorous approach by validating the outcome variable for all potential cases and non-cases to accurately classify the outcome variable for each subject. Overall, the required extent of validation should be determined by*

necessary level of certainty and the implication of potential misclassification on study inference."

Given the relationship among PPV, NPV, sensitivity, and specificity, we suggest moving this text to above Table 1 (line 496) or moving Table 1, accompanying text, and the text in lines 853-866 to an appendix.

* * *

Aetion looks forward to collaborating with the Agency to support the successful use of RWD and RWE in regulatory decision-making. Please contact Lowell Schiller at Lowell.Schiller@Aetion.com with any questions regarding these comments or other issues related to RWE policy and development.

Sincerely,

DocuSigned by:
Lowell Schiller
377634F47A5048F...

Lowell Schiller
Chief Legal and Regulatory Officer, Aetion