AETION

February 4, 2022

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

Re: Docket No. FDA-2021-D-0548 "Data Standards for Drug and Biological Product Submissions Containing Real-World Data; Draft Guidance for Industry"

To the Food and Drug Administration:

Thank you for publishing the draft guidance document entitled "Data Standards for Drug and Biological Product Submissions Containing Real-World Data; Draft Guidance for Industry" (FDA-2021-D-0548) (the "Draft Guidance").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 and enhance understanding of product safety and effectiveness. 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Aetion welcomes and appreciates the opportunity to comment on the Draft Guidance.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 technology 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 and effectiveness of clinical interventions. A key feature of AEP is data fluency, meaning that it enables users to run analyses against different data sources in their near-native formats, with minimal transformations required prior to analysis.

Aetion shares FDA's commitment to fostering high-quality analyses of fit-for-purpose 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. Greater alignment on principles and best practices, and greater public visibility regarding FDA's expectations, will enable sponsors and other researchers to implement RWD analyses in a transparent, auditable, reproducible, and scientifically valid manner.

**Comments on the Draft Guidance**

We appreciate FDA's work to provide guidance on how Section 745A of the Federal Food, Drug, and Cosmetic Act applies to RWD submitted in applicable drug and biological product submissions. Section 745A, which authorizes FDA to require that certain electronic submissions to the Agency be submitted in a standardized and pre-specified format,

provides a powerful tool to help FDA ensure that the electronic data it receives as part of a submission are in a format that the Agency can support.

We agree with FDA that when a submission contains RWD, it is important that the data be submitted to the Agency in a usable data model and format. Reviewers should be able to characterize, evaluate, and re-analyze the data efficiently, and the data format should facilitate, rather than impede, the Agency's ability to determine whether the data in an RWE study are reliable and relevant, and whether the results are reproducible. First and foremost, however, we believe that selecting a data source for a particular research question should be based on scientific considerations. Blanket formatting requirements may lead to a selection of a data source based on ease of meeting those requirements, rather than what dataset is most fit for purpose. Additionally, requiring conversion of data from their native format to a particular data standard may have significant implications for a study's ability to meet the transparency and traceability recommendations that FDA has put forward in other draft guidance, such as the "Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products" (September 2021) ("Draft Claims/EHR Guidance").

Unlike data collected in a clinical trial, RWD often is collected in "secondary use" settings where the primary purpose is not to generate evidence to support a regulatory decision by FDA, but rather (for example) to support clinical workflows or the payment of an insurance claim. As a result, data collected from real-world settings come from a wide variety of sources and are stored in a wide variety of models and formats.

Given this variety of data models and formats, we appreciate that greater interoperability and standardized data collection early in the data life cycle would improve the efficiency of conducting and reviewing RWE studies in certain respects. However, we believe that the approach proposed in the Draft Guidance will undermine the efforts to generate high-quality RWE more than it will assist. We urge the Agency to recognize that sponsors may be able to use alternative approaches, such as a validated RWE platform, that are equally permissible under law and that achieve the same objective — *i.e.,* ensuring that the Agency can efficiently review and replicate analyses involving RWD — with less disruption to the generation of high-quality RWE and greater impact on the public health.

1. *Requiring RWD in submissions to be transformed into a standardized data format would have material negative impacts on study quality and availability.*

We believe a requirement to transform RWD submissions into a standardized format would be disruptive to ongoing efforts to generate high-quality RWE for use in regulatory decision-making. For FDA to be able to independently verify and replicate the results of a submitted study — as the Agency would with clinical trial data — sponsors would look to analyze the RWD in the format in which it would be required to be submitted. This means that if FDA were to require study data to be submitted in a standardized format, the sponsor would likely transform any RWD into that format as an early step of the study process and likely do

so without benefit of the context of how the data will be used in the study (*e.g.*, before all potential confounders are identified).

As a result, the practical effect of the Draft Guidance, if finalized, would be to narrowly limit the data formats that may be used for studies that may be submitted to FDA, and to narrowly limit the formats in which data are stored and analyzed. While data analytics companies like Aetion have the capability to help sponsors implement such a requirement, and might even stand to benefit from some of the effects of standardization, doing so would have significant implications for study quality and timing. This is acknowledged in the Draft Guidance, which identifies multiple "challenges involved in standardizing study data derived from RWD sources for inclusion in applicable drug submissions." (Lines 85-86.) Based on our experience, we believe the challenges and associated drawbacks would be more substantial than what is described in the Draft Guidance — in many cases, to the point that the negative consequences (four of which are described below) would substantially outweigh the benefits of standardization.

*First*, although converting study data into a supported format may be intended to create efficiencies for both researchers and FDA reviewers, the impact could diminish study validity. A requirement to transform RWD to a supported data standard is, in practical effect, a mandate to use a particular common data model (CDM) — which FDA has described as "[s]tandardiz[ing] a variety of electronic health care data sources into a common format to ensure interoperability across all sites providing data." (Draft Claims/EHR Guidance, lines 1259-1260). In the Draft Claims/EHR Guidance, FDA recognized (at lines 329-354) that "CDMs can introduce additional challenges for researchers to consider" and recommended that sponsors should assess CDMs "to determine suitability for the study and whether identified deficiencies can be addressed." We agree with this recommendation: an important component of fit-for-purpose data selection is determining whether the data format is sufficiently amenable to a CDM, especially if linkage among multiple data sources is required. The Draft Guidance, however, would effectively take the formatting decision away from the researcher and prescribe a particular format, regardless of whether it is the right fit for the study in question.

Currently, the only relevant standards supported in FDA's data standards catalog are Clinical Data Interchange Standards Consortium's (CDISC's) Study Data Tabulation Model (SDTM) and Analysis Data Model (ADaM) formats. These formats were designed for clinical trials and are not optimized for RWE. Moreover, unlike some CDMs which are designed to facilitate queries across datasets while preserving the underlying source data, CDISC's SDTM and ADaM formats fall into the category of "mapping" CDMs (sometimes referred to as preconfigured rules systems) — *i.e.*, CDMs that map the data onto a standardized set of constructs, thus potentially modifying the intended meaning of the source data. Whereas "organizing" CDMs 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, as has been discussed in peer-reviewed publications.[1]

In practice, RWD sources typically do not contain enough information to appropriately map RWD to CDISC SDTM or ADaM format. For instance, raw variables, data collection practices, and underlying algorithms used to curate data may not be readily apparent or easily quantified, which could impact the fidelity of the mapping process.[2] The lack of such information would need to be addressed *a priori* at the time of study design, and the ability to continue with the dataset would need to be verified in feasibility analyses; undertaking a study without full context could unintentionally bias study results.[3] While a potential workaround might involve working with a data vendor to include all necessary information to permit complete mapping, this ability is subject to data disclosure agreements and could raise concerns around patient privacy.[4]

Additionally, existing SDTM and ADaM concepts do not entirely account for the nuances of the data collected in observational studies.[5] As such, there are a number of situations where ADaM would be an unworkable data standard and use of SDTM might result in a situation where SDTM rules are more often violated than fulfilled.[6] If the Draft Guidance is finalized to require mapping to preconfigured rules, it will become much more challenging — and perhaps impossible in some cases — for a study to adhere to the recommendations of the Draft Claims/EHR Guidance to ensure transparency across the data life cycle.

*Second*, mapping RWD to CDISC SDTM or ADaM may have other unforeseen impacts on the validity of RWE studies. In the Draft Guidance's Appendix, the Draft Guidance recognizes that the exercise of mapping to CDISC SDTM or ADaM is a complex task, particularly when a high level of detail needs to be preserved. (*See, e.g.*, lines 321-333). The complexity of the task introduces new opportunities for human error (and need for FDA reviewer verification) that would not exist if data were analyzed in near-native format. Moreover, even absent human error, mapping to CDISC SDTM or ADaM can result in the loss of critical detail, as described in the Appendix to the Draft Guidance. When detail from a dataset is lost, it can materially affect the validity of a study or the fitness of the dataset for answering a particular research question. Indeed, a recent survey conducted by the CDISC RWD Connect Initiative concluded

---

[1] Schneeweiss, S., Brown, J.S., Bate, A., Trifirò, G. and Bartels, D.B. (2020), Choosing Among Common Data Models for Real-World Data Analyses Fit for Making Decisions About the Effectiveness of Medical Products. Clin. Pharmacol. Ther., 107: 827-833. https://doi.org/10.1002/cpt.1577; Seesaghur, A, Petruski-Ivleva, N, Banks, V, et al. Real-world reproducibility study characterizing patients newly diagnosed with multiple myeloma using Clinical Practice Research Datalink, a UK-based electronic health records database. Pharmacoepidemiol Drug Saf. 2021; 30: 248– 256. https://doi.org/10.1002/pds.5171.

[2] Thorne N, Wu K, Sundaram R. Real World Evidence at Janssen. PhUSE USA Connect 2020. Available at: https://phuse.s3.eu-central-1.amazonaws.com/Archive/2020/Connect/US/Virtual/PAP_SA03.pdf and https://phuse.s3.eu-central-1.amazonaws.com/Archive/2020/Connect/US/Virtual/PRE_SA03.pdf.

[3] *Id.*

[4] *Id.*

[5] Bahatska Y and Ivanushkin V. First Aid Kit for an Observational Study. PhUSE EU Connect 2018. Available at: https://phuse.s3.eu-central-1.amazonaws.com/Archive/2018/Connect/EU/Frankfurt/PAP_PP24.pdf.

[6] *Id.*

that "Data elements related to longitudinal, prospective, and observational study designs are not sufficiently modeled in CDISC standards currently."[7]

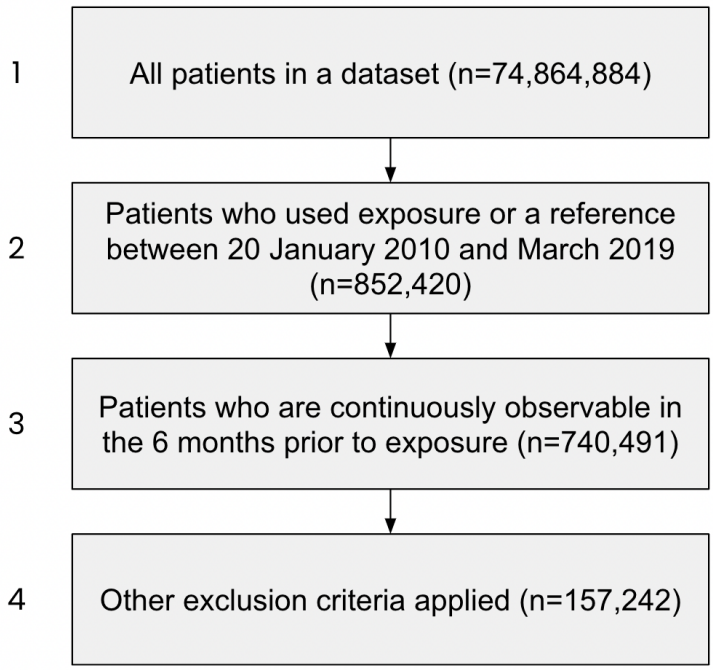
While techniques for preserving detail exist, they can be "labor-intensive" (as the Draft Guidance acknowledges at line 324) and imperfect. One example of such imperfection is the effect that mapping to CDISC SDTM can have on a study's ability to account for patients' observability in a dataset[8] — that is, when a patient's health data would be expected to be captured in the dataset and when it would not be — which helps provide the denominator in an RWE study. When utilizing RWD that is not based on electronic case reports (eCR), such as administrative claims or electronic health record (EHR) derived datasets, the ability to account for observability is a critical component to robust study design often used in cohort definitions and outcome ascertainment. Some common data models, such as Sentinel, have built-in mapping variables or modules to account for capturing observability. But because CDISC SDTM was designed for clinical trial data — where patients are generally assumed to be observable through the course of the trial — this standard does not have similar mapping variables to account for observability. While workarounds exist for mapping observability into CDISC SDTM, they are time-consuming and resource intensive, and may require making assumptions about observable and non-observable person-time that ultimately reduce the quality of the resulting RWE.

Continuing this example, exclusion criteria in real-world studies often require a minimum baseline period of observability to uniformly capture confounding variables and otherwise mitigate potential sources of bias. In Figure 1 below, step 3 has been implemented to introduce an inclusion criterion that accounts for observability. In CDISC SDTM format, this critical step may not be possible through an available mapping variable, and may need to either be omitted or achieved through assumption rather than direct measurement.

[7] Facile R., Muhlbradt EE., Gong M., Li Q., Popat V., Pétavy F., Cornet R., Ruan Y., Koide D., Saito TI., Hume S., Rockhold F., Bao W., Dubman S., Jauregui Wurst B. (2022), Use of Clinical Data Interchange Standards Consortium (CDISC) Standards for Real-world Data: Expert Perspectives From a Qualitative Delphi Survey JMIR Med Inform 2022;10(1):e30363. doi: 10.2196/30363.

[8] As used here, "observability" refers to the time during which the patient is active in the database and, if the patient were to have an event of interest, that event would be recorded.

Figure 1: Flow Chart Adapted from RCT-Duplicate[9]




*Third*, the complexity inherent in converting data into a standardized format like CDISC SDTM or ADaM is likely to be exacerbated in studies that involve linking disparate data. Although the Draft Claims/EHR Guidance provides helpful considerations regarding linking data, it is not clear how that guidance should be applied in light of statements in the Draft Guidance that study data should be converted to a supported format. If multiple levels of transformation are required to create a single, analyzable linked dataset, the added complexity and ambiguity may be great enough to discourage some studies that might otherwise be undertaken, or prompt sponsors to opt for different, and perhaps less robust, study designs than they otherwise would. This would be a loss; data standardization should not be a limit to appropriate study design.

*Fourth*, because mapping data to CDISC SDTM or ADaM while preserving detail is likely to be "labor-intensive" (Draft Guidance line 324), requiring this transformation, rather than the use of a near-native data format, is likely to introduce significant delay in the time it takes to execute a study involving RWD. Mapping data to a CDM, and especially CDISC SDTM or ADaM, is difficult to automate and requires extensive manual oversight and effort. One published study showed that only 55% of its data elements mapped to core SDTM tables, with 45% of the

[9] Franklin, J., 2022. [online] Clinicaltrials.gov. Available at: <https://www.clinicaltrials.gov/ProvidedDocs/49/NCT03936049/Prot_SAP_002.pdf> [Accessed 17 January 2022].

data requiring manual manipulation, thus leading the researchers to conclude that SDTM "significantly increases burden at the time of data use."[10]

A key objective of the 21st Century Cures Act, which required FDA's RWE Framework, is to "accelerate medical product development and bring innovations faster and more efficiently to the patients who need them."[11] The additional time and effort associated with mapping to a standard format would have material impacts on that goal, particularly in the context of a public health emergency like the COVID-19 pandemic or therapeutic areas with rapidly changing standards of care, such as oncology. Studies involving RWD often are executed in contexts where real-time insights are critical, and adding this extra step will impede FDA's access to time-sensitive information that may bear upon the public health.

2. *FDA can support the review of RWD in diverse formats through means that require minimal transformations, increasing study validity.*

We share FDA's goals of promoting the efficiency of product reviews and ensuring that the Agency has the technical capability to support the review of the data submitted to it. However, we believe that requiring data conversion to a supported standard is not necessary to support these goals. To the contrary, other tools are available that enable FDA to efficiently review studies conducted using data in their near-native formats with minimal transformations, thereby avoiding the challenges associated with conversion to a required format and increasing study precision and validity.

Accordingly, we believe the Agency should consider alternate approaches to standardizing the review of RWD submitted to support regulatory decision making. For example, technologies such as centralized, platform-based analytics enable the review and analyses of RWD from a variety of sources, without converting to a single supported data standard. Such an approach could enable reviewers to access to source data in a transparent manner while maintaining information and detail that might be lost in the process of mapping the data to a particular standard.

An RWE platform can provide fully transparent documentation of every methodological decision made over the course of the analysis. It can document data provenance and data transformation, store data long-term, provide access to appropriate parties, and document data linkage and analytic cohorts. In addition, a platform provides sponsors the ability to transmit and exchange data containing proprietary information without jeopardizing the confidentiality of that information. Finally, FDA reviewers can review studies on the platform, fully querying RWD in any model or format without the need to interact directly with patient-level records, removing the need for standardization into a supported data standard.

---

[10] Garza M, Del Fiol G, Tenenbaum J, Walden A, Zozus MN. Evaluating common data models for use with a longitudinal community registry. J Biomed Inform. 2016;64:333-341. doi:10.1016/j.jbi.2016.10.016.
[11] Food and Drug Admin, Framework for FDA's Real-World Evidence Program. Available at: https://www.fda.gov/media/120060/download.

Should further re-analysis be desired, platforms also facilitate the download of data in simple, easy-to-analyze formats such as comma-separated values (CSV).[12]

Section 745A leaves ample room for FDA to accept such alternative approaches, and we encourage the Agency to do so. With the technology available today, tools exist for FDA to efficiently review studies with minimally transformed data, without the negative unintended consequences associated with requiring across-the-board data standardization. We encourage FDA to fully leverage available technology to avoid such unintended consequences.

*  *  *

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

[12] Other alternative approaches may be feasible, as well. For example, another approach that would facilitate efficient review while preserving flexibility in formatting would be submitting RWD together with R scripts that FDA reviewers could use to validate or even replicate a submitted study. There are likely other options that would facilitate the same goals.