Guardian Research Network

101 East Wood St.
Spartanburg, SC 29303
guardianresearch.org
864.777.7165

November 29, 2021

Docket Management Staff (HFA-305)
Food and Drug Administration
5630 Fishers Lane, Room 1061
Rockville, MD 20852

Subject: 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

Dear Sir or Madam:

Guardian Research Network, Inc. (GRN) is a non-profit, tax-exempt corporation organized and operated under Section 501(c)(3) of the Internal Revenue Code based in Spartanburg, SC. Enhancing clinical trial quality and improving patient outcomes involves interactions among healthcare providers, researchers and service providers sharing a common interest in expanding both the quality of clinical care coordination as well as the quantity of treatment alternatives. Ongoing challenges and inefficiencies in clinical trial enrollment stem largely from the realization that a large percentage of eligible patients are geographically and/or socioeconomically unable to receive treatment where the best clinical trials are being offered. For such marginalized patient populations, enrolling in a clinical trial often requires the patient to expend significant time, cost, and energy in traveling to the study site.

We greatly appreciate the opportunity to provide our comments in response to the draft guidance, *Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision Making for Drug and Biological Products* issued September 30th, 2021. This draft guidance provides sponsors, researchers, and other interested stakeholders with considerations when proposing to use electronic health records (EHRs) or medical claims data in clinical studies to support a regulatory decision for effectiveness or safety.

As the Agency continues to develop this guidance, GRN respectfully requests the FDA consider the commentary raised in this submission. Future revisions to this guidance document should stress the dynamic nature of this topic and highlight FDA's flexibility and receptivity in engaging with data providers to creatively explore other concerns, approaches, and considerations specific to drug development programs.

Guardian Research Network thanks FDA for valuable insight and guidance on this topic.

Mark L. Watson

Mark Watson, M.D., Ph.D.
Chief Operating Officer
Guardian Research Network

R. Bakin

Robert Bakin, Ph.D.
Chief Compliance & Privacy Officer
Guardian Research Network

101 East Wood St.
Spartanburg, SC 29303
guardianresearch.org
864.777.7165

General Commentary

The use and disclosure of RWD/RWE in support of regulatory decision making is a rapidly evolving space. While all stakeholders should strive to curate high quality, reliable and relevant EHR data, great care must also be taken to ensure patient privacy. Patient privacy must never be sacrificed at the expense of broader use of EHR data for regulatory decision making. Strong, HIPAA-compliant data security measures must always be in place to ensure the responsible use and disclosure of EHR data for all purposes. ***Patient privacy comes first.***

Enhancing clinical trial quality and improving patient enrollment involves interactions among healthcare providers, researchers and service providers sharing a common interest in expanding both the quality of clinical care coordination as well as the quantity of treatment alternatives. Regulators are increasingly concerned about the underrepresentation of minorities from industry sponsored trials. Acknowledging a general theme of FDA's September 2021 draft Guidance for Industry "*Real-World Data: Assessing Electronic Health Records and Medical Claims Data To Support Regulatory Decision-Making for Drug and Biological Products*", we note that community-based hospital systems offer a unique opportunity to positively impact patient populations having extensive continuity of coverage because of socioeconomic factors. ***All patients matter.***

While acknowledging that the sponsor is ultimately accountable, we recommend FDA clarify the role of the data provider (e.g., EHR data curator/vendor) in support of regulatory decision making. As part of the review of study feasibility, we recommend that sponsors obtain from the data provider evidence of the completeness, accuracy, and plausibility of the data as well as the scientific justification for using data standards. ***Demand data accountability.***

Specific Commentary

Section IV. DATA SOURCES: General Discussion *(line 190)*

Concern: GRN clientele have often expressed concern that EHR data collected during routine care often lacks common prespecified research protocol inclusion/exclusion criteria.

Recommendation: For community healthcare systems actively involved (or seeking to become involved) in clinical drug trials, we seek specific FDA commentary on addition of the following clinical trial criteria to routine care EHR: (i) patient performance status (e.g., ECOG), (ii) patient therapy response and date of disease progression, (iii) common acute and chronic conditions that often contraindicate participation (e.g., New York Heart Association Class IV CHF), and (iv) vaccination records. We acknowledge the potential for additional burden on community healthcare workers to provide such critical inclusion/exclusion criteria and/or participate in additional training.

Section IV. DATA SOURCES: Enrollment and Comprehensive Capture of Care *(line 201)*

Concern: Continuity of patient coverage is frequently lacking in geographic areas having many healthcare options (e.g., urban areas, major cities, etc.), leading to biases/gaps/inconsistencies in source data.

Recommendation: Relative to urban settings, continuity of coverage is generally much higher in rural, socioeconomically underserved healthcare systems. As such, "rural" EHR data may have less bias than from regions of the county with multiple healthcare provider options. If concerns over continuity of coverage would likely affect clinical readouts, FDA may want to consider emphasizing

101 East Wood St.
Spartanburg, SC 29303
guardianresearch.org
864.777.7165

rural patient enrollment. Similarly, data providers who curate EMR data from a variety of healthcare systems *within a geographic region* (urban or rural) are often able to extend continuity of patient coverage by merging EHR data from distinct healthcare providers who are members of a consortium of data providers (e.g., Guardian Research Network ®).

Section IV. DATA SOURCES: Data Linkage and Synthesis *(line 248)*

Concern: Clinical studies requiring the combination of data from multiple sources (e.g., disparate healthcare providers, data vendors, disease registries, etc.) run the risk of inclusion of duplicate data for a single individual (or loss of redundant data from a single individual due to death), thus introducing bias into data readouts. This is particularly problematic when patient identities have been de-identified to federal HIPAA standards.

Recommendation: (i) FDA consideration of an anonymous "results-based checksum" capable of identifying and eliminating redundant de-identified sources of EMR data. Such a checksum would be based on statistical analysis of relevant standardized clinical metrics like diagnosis and procedure codes (e.g., ICD-9-CM, ICD-10-CM) and laboratory tests (e.g., LOINC) and (ii) FDA consideration of requiring data vendors routinely communicate to recipients of curated data whether the source data came from a single healthcare network or multiple networks.

Section IV. DATA SOURCES: Unstructured Data *(line 369)*

Concern: Valuable clinical metadata exist as unstructured data within EMRs and are lost during data aggregation, curation, and de-identification.

Recommendation: GRN recommends: (i) data vendors continually involve data engineering teams to ensure maximal extraction of unstructured data elements from EHRs, (ii) human review of all curated data sets prior to use, particularly for data sets created by automated means (e.g., NLP, AI, etc.), (iii) not relying only on standard data export formats (e.g., FHIR, HL7, etc.) that often lack important clinical metadata (e.g., physician notes), (iv) clinical protocols 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, and (v) FDA reviewable QA/QC validation reports from data vendors who employ automated software means of data curation (akin to November 2021 FDA guidance on "*Content of Premarket Submissions for Device Software Functions")*.

Section IV. DATA SOURCES: Missing Data *(General Considerations, line 395)*

Concern: Missing and/or misclassified information can hinder effective use of EHR data for regulatory decision-making.

Recommendation: FDA provide more accurate definitions for otherwise vague routine clinical care terminology (e.g., more precisely define "date of diagnosis" as date of pathology lab diagnosis / date of clinical lab values diagnosis / date patient was informed of diagnosis).

Section V. STUDY DESIGN ELEMENTS: Selection of Study Population *(line 538)*

Concern: Often, GRN will identify a surplus of patients for clinical trials who meet all enrollment criteria. Recommending final patient cohorts can be challenging because of selection bias.

101 East Wood St.
Spartanburg, SC 29303
guardianresearch.org
864.777.7165

Recommendation: FDA work with stakeholders on improving propensity score matching by identifying key RWD covariates. Data vendors should display full transparency by providing the entire patient cohort and describing how final patient cohorts were selected in an unbiased fashion.

Section VI. VI. DATA QUALITY DURING DATA ACCRUAL, CURATION, AND TRANSFORMATION INTO THE FINAL STUDY-SPECIFIC DATASET *(Characterizing Data, Data Accrual, line 1090)*

Concern: Methods for data retrieval and processes to minimize implausible values captured at the point of care (e.g., during clinical practice for manual or automated health care data collection) occasionally eliminate accurate values and weakening data integrity.

Recommendation: To best maintain data integrity, instead of deleting or "correcting" seemingly improbable values, GRN recommends documenting and flagging suspicious data in need of further review and confirmation by sponsors. GRN has first-hand knowledge of seemingly implausible clinical values (e.g., WBC count) ultimately being proven accurate during curation QA/QC process.

Section VI. VI. DATA QUALITY DURING DATA ACCRUAL, CURATION, AND TRANSFORMATION INTO THE FINAL STUDY-SPECIFIC DATASET *(Characterizing Data, Data Transformation, line 1154)*

Concern 1: While accrual and curation of quality EHR data is essential for regulatory decision making, patient privacy issues must be contemplated during data transformation. We note that use of certain automated EHR de-identification algorithms (e.g., "lossy" algos) are occasionally used to ensure data anonymization and prevent re-identification. Unfortunately, such algorithms are quite susceptible to brute force patient re-identification when trained on a large database of random patient identifiers.

Recommendation: Parties who use and disclose identifiable EHR data for regulatory decision-making purposes should be legally bound to HIPAA regulations via Business Associate Agreements / Data Use Agreements. Non-HIPAA contractual obligations, while useful, are insufficient to ensure patient privacy and prevent improper third-party re-identification.

Concern 2: Data transformations using automated de-identification algorithms (e.g., NLP) may be prone to improperly retaining protected health information ("false negative") or improperly deleting non-protected health information ('false positive").

Recommendation: Data providers should: (i) establish policies and procedures for developing and testing the de-identification process, and (ii) routinely create validation reports prior to electronic transmission to third-party end users.

Section VI. VI. DATA QUALITY DURING DATA ACCRUAL, CURATION, AND TRANSFORMATION INTO THE FINAL STUDY-SPECIFIC DATASET *(Documentation of the QA/QC Plan, line 1218)*

GRN agrees that a robust QA/QC multidisciplinary approach including clinical input is necessary to ensure adequate capture and handling of data, particularly for EHR systems, which inherently incorporate nuances and intricacies of health care delivery. GRN recommends the use of a dedicated, HIPAA compliant clinical operations group to oversee QA/QC issues.

Section VI. VI. DATA QUALITY DURING DATA ACCRUAL, CURATION, AND TRANSFORMATION INTO THE FINAL STUDY-SPECIFIC DATASET *(Documentation of the Data Management Process, line 1232)*

101 East Wood St.
Spartanburg, SC 29303
guardianresearch.org
864.777.7165

GRN agrees that all manual and automated data retrieval and transformation processes should be thoroughly assessed from data collection through writing of the final study report to ensure data integrity. Descriptions of processes should include safeguards or checks to ensure that patient data are not duplicated or overrepresented. GRN recommends the use of a dedicated clinical operations group to oversee documentation of data management issues.