



A driving force for health equity

Submitted via [webform](#)

December 14, 2023

Dr. Susan K. Gregurick
Director
Office of Data Science Strategy
National Institutes of Health
9000 Rockville Pike
Bethesda, Maryland 20892

Re: Inviting Comments and Suggestions on Opportunities and Challenges for the Collection, Use, and Sharing of Real-World Data (RWD) including Electronic Health Records, for NIH Supported Biomedical and Behavioral Research (NOT-OD-23-180)

Dear Dr. Gregurick,

On behalf of OCHIN, we welcome the opportunity to provide responses to the topics listed in the Request for Information *on Opportunities and Challenges for the Collection, Use, and Sharing of Real-World Data (RWD) including Electronic Health Records, for NIH Supported Biomedical and Behavioral Research*.

Background

OCHIN is a national nonprofit health information technology and research network that offers technology solutions, informatics, evidence-based research, and policy insights. OCHIN serves nearly 2,000 community health care sites with 25,000 providers in 40 states, reaching more than 8 million patients. For over two decades, OCHIN has advanced health care solutions by leveraging the strength of our network's unique data set and the practical experience of our members to drive technology innovation for patients and providers in rural and other underserved communities. Our network supports high-quality care for underserved and under-represented groups impacted by health disparities.

In addition to supporting the largest network of federally qualified health centers (FQHCs) and other community clinics, OCHIN supports certified community behavioral health clinics (CCBHC), complex specialty mental health organizations, local public health agencies, corrections, school-based mental health and substance use disorders programs, and youth authorities. OCHIN's members serve rural, partially rural, urban, and suburban communities. Nearly 30% of the network's patients are best served in a language other than English, over 40% are Persons of Color and nearly 26% are Hispanic/Latino. In addition, 23% are children, almost 11% are seniors and 2% are veterans. Among the patients who received care via an OCHIN supported electronic health record (EHR) platform in 2022, nearly 1 out of 2 are covered by Medicaid, and another quarter were uninsured; nearly 1 out of 2 had two or more chronic conditions; and more than half of our network patients live at or below the federal poverty level. OCHIN has also led efforts to build one of nine PCORnet clinical research networks funded by Patient Centered Outcomes Research Institute (PCORI) and OCHIN's researchers are national leaders in health equity research. The OCHIN led PCORnet network is the nation's largest community health center

focused research networks and is the only one that includes data on underinsured and uninsured patients.

Scientific value and quality considerations for collection, use, and sharing of RWD in biomedical and behavioral research.

The scientific value of real-world data (RWD), including EHR data, is well recognized. However, RWD, particularly EHR data, is not originally collected for research purposes, which raises a series of issues to consider. Because of the multiple sources of bias associated with the generation, storage, and extraction of RWD, **we recommend that standards for access to RWD include constraints for its use that mitigate bias and preserve scientific rigor.** Examples of sources of bias in RWD and its use include the following. The inclusion of individual patient information hinges on factors such as access to and utilization of healthcare facilities. Because social determinants of health are determinants of access to healthcare, healthcare utilization patterns, baseline health status – including chronic conditions – how clinicians document care, how clinicians make medical decisions, and patient outcomes, the quantity and quality of EHR data is, itself, dependent on these social determinants. The presence or absence of specific observations documented within EHRs depends on a multitude of factors. This includes clinician and patient decision-making, documentation practices, insurance and billing processes, illness severity, and other contextual factors. In addition, structural racism can introduce inherent biases in how patients interact with the healthcare system, and failing to address these biases could perpetuate disparities in subsequent analyses using these data. Therefore, it's imperative to acknowledge that for researchers working with EHR data, collaborative efforts with those responsible for data creation and extraction are essential to provide context to the eventual analytic dataset. Collaboration with those well-versed in the clinical context of the EHR system on the subjective process of extraction, customization, and operationalization is critical for the responsible conduct of research with these data. **Consequently, all use and sharing of EHR data should be firmly rooted in scientific collaboration with those who understand the context in which this data was originally generated and supported by other approaches (e.g., meta data with clear field information).**

Using RWD as part of the scientific paradigm, including open science, scientific rigor and reproducibility, and team science.

It is essential to recognize that RWD, particularly EHR data, face significant obstacles to participation in open science and collaborative team science, posing threats to reproducibility and scientific rigor. In the evolving landscape of scientific research – where there is a growing emphasis on open science, scientific rigor and reproducibility, and collaborative team science – we recognize that RWD, particularly EHR data, often do not fit as easily with biomedical approaches to these shifts and cannot be shared. First, sharing data in open science presents a risk of re-identifiability and harm to those whose data are included in the study – a risk that is only magnified when appreciating that many patients are likely not fully aware of how their data are being used. **We recommend continued attention to the risks of data sharing in this context and clear guidance for RWD in required documents (e.g., in Data Management and Sharing Plans) and clear uniform parameters around the circumstances in which data could be shared and in what format (e.g., Data Use Agreement).** Second, there are unique issues with RWD when considering reproducibility. EHR data are dynamic and variable across clinical settings and over time. This severely limits the true reproducibility of scientific inquiry. As an example, if researchers want to pursue the same research question/phenomenon in the same clinical setting but at a later point in time, they cannot rely on previously written code for data extraction and analysis without first verifying that the same code sets (e.g., ICD/CPT), EHR functionality, and clinical data collection processes are

being used. Third, the operationalization of key study variables needs tailoring and evolution to accurately represent patients' health and healthcare utilization within their medical and community context. **Therefore, while the principles of open science and reproducibility are important it is also critical to recognize that this is not achieved by the simple application of published analyses to data.**

Administrative and logistical considerations for collecting, using, and sharing RWD for biomedical research.

The linkage and sharing of RWD, especially EHR data, present exciting opportunities, thanks to tokenization and privacy-preserving advancements. These technologies hold the potential to address privacy concerns and make data more accessible. However, responsible realization of these advances will demand significant administrative and financial resources. **First, infrastructure support is crucial to implement and maintain these innovations effectively.** Administrative processes such as Institutional Review Board (IRB) approvals remain challenging and time-consuming to navigate while ensuring data use aligns with ethical and legal standards. And the practicality of seamlessly integrating data from all these different sources is often limited—and in many cases, simply not feasible— due to differences in data structures, formats, and governance.

Second, metadata derived from EHRs must be clearly and specifically communicated. EHR data may contain internal biases due to the setting that constrains what data are medically relevant and therefore included in the EHR. Informed by this clinical and community context, EHR researchers often must "compose" data elements by piecing together relevant components from the EHR. This process can involve accumulating disparate observations that represent a particular concept and consolidating them into data that is suitable for analysis; essentially curating datasets in an "artisanal" sense.

From a health equity perspective, the use of RWD from populations who experience health disparities must be taken with even more caution. They have often experienced historical and current injustices in biomedical research and the protection and preservation of their privacy remains of utmost importance. **Given this, focus should not be placed on the speed of data access but rather the partnership process in place to ensure that access is only granted when regulatory approvals are in place and in the context of a relationship of trust between the internal research team who have detailed knowledge of the EHR data and external partners wanting to use that data.** From a scientific point of view, this relationship of trust helps ensure that complexities of these data elements related to populations who experience health disparities are understood theoretically and analytically. For instance, it is difficult to know whether race and ethnicity data are self-reported by patients versus presumed by clinical staff members; patients may or may not have the ability to select more than one racial/ethnic category; and available racial/ethnic categories may not align with any of a patient's identities. **Framing RWD as a commodity for purchase and promoting swift access has the potential to undermine not just the critical scientific merits but also the trust of the individuals for whom these data represent.** Therefore, those working with RWD should consider the subpopulations in their data who experience health disparities or historical injustices in research and consider the implications of this in the collection, use and sharing of the data.

Ethical considerations for using RWD for biomedical and behavioral research.

There are several ethical considerations in the use of RWD. **One of the foremost is the need to remain cognizant that meaningful informed consent remains a work in progress and many patients are likely not fully aware how their personal health information is being used in scientific inquiry; even if they**

have signed legal waivers to generally allow its use. With the history of harm that biomedical research has inflicted, we must not perpetuate this by using RWD in a manner that could do so. As the attention to 'big data' and complex quantitative approaches has only grown, we must strive to meaningfully engage patients and communities in the use of data that represent them. We strongly urge NIH to support development of consent methods and processes that are grounded in user-centered design principles and reflect the conditions and needs of populations experiencing health disparities. **Further, another strategy is the creation of community or patient engagement panels, which allows for formal scientific partnership and oversight from individuals whose data are involved in research.** We encourage the employment of this strategy and the development of additional strategies to guard against potential harms of using RWD in research. For example, OCHIN is proud to work with our Patient Engagement Panel (PEP) which serves not as qualitative study participants, but rather as scientific partners and advisors to OCHIN research studies.

If you have questions, please contact Wyatt Bensken, PhD, Research Investigator, benskenW@ochin.org. We welcome the opportunity to elaborate further on the recommendation contained in this letter.

Sincerely,

A handwritten signature in blue ink that reads "Jennifer Stoll". The signature is fluid and cursive, with the first name being more prominent.

Jennifer Stoll
Executive Vice President
OCHIN