Duke | MARGOLIS CENTER for Health Policy

Establishing a High-Quality Real-World Data Ecosystem

Duke-Margolis Center for Health Policy | 2-Day Online Workshop

Monday, July 13 | 1:00 – 4:00 pm ET

Tuesday, July 14 | 1:00 – 4:30 pm ET

Annotated Agenda

Working Toward High-Quality Real-World Data

The systematic collection of high-quality data is critically important for effective health care systems. Data regarding patients' health and health care delivery is routinely recorded to support both clinical decisions and payment for services. Examples include clinical and lab data, insurance claims, patient-generated health data as well as socio-economic, environmental, genomic, and other emerging types of data sources. Stakeholders across the health system have identified opportunities to leverage such real-world data (RWD) for broader evidence generation needs, such as exploring the effectiveness and efficiency of health systems, supporting public health practice, and accelerating innovative clinical research with medical products.[1]

In 2018, the U.S. Food and Drug Administration published a strategic framework to guide the development of a new regulatory program to leverage RWD and resultant real-world evidence (RWE) for regulatory uses, such as supporting changes to medical product labeling.[2] The framework uses a three-pronged approach on how the Agency would evaluate RWE, which starts with a fit-for-purpose dataset that is reliable and relevant to the regulatory question of interest.

Determining whether RWD are fit-for-purpose is a complex and multifaceted process.[3] Existing RWD sources often do not adequately or consistently capture the types of health outcomes and endpoints important for clinical research. Because the data are not generated for research purposes, there can be substantial diversity or differences in data capture practices within and between RWD sources. There is also variability in how these data are transformed and curated from source inputs into a fit-for-purpose dataset that is ready for analysis. Having insight into the original data collection procedures, provenance, and subsequent data transformations is critical to fully understanding the strengths, limitations, and quality of the data.

To date, efforts to improve data quality have focused on data curation practices and developing standardized approaches for checking data quality. Although data curation is a fundamental component of developing RWD for regulatory use, key challenges remain with respect to the time, resources, and transparency needed to monitor and track data transformations as well as the scalability of these efforts especially when utilizing large RWD databases over long periods of time.

Another area of interest has focused on performing checks of data reliability to ensure it adequately represents the medical concepts intended. Distributed data networks, which utilize a common data infrastructure and data model across participating health care sites, have designed processes for

[1] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2329823/
[2] https://www.fda.gov/media/120060/download
[3] https://healthpolicy.duke.edu/sites/default/files/2020-03/characterizing_rwd.pdf

reviewing and verifying the quality of data in the network. Noteworthy examples can be drawn from the Sentinel Initiative, The National Patient-Centered Clinical Research Network (PCORnet), and the Observational Health Data Sciences and Informatics (OHDSI) Collaborative.[4,5] However, documenting curation practices, appropriate data model mappings, and ensuring accurate characterizations of data quality through these checks also requires substantial time, resources, and human judgement. Furthermore, such checks may also impose practical challenges on regulators who must review the documentation of these checks when evaluating whether a submitted dataset meets fit-for-purpose requirements.

Given these challenges with data quality, there is growing stakeholder awareness that improving data capture practices at the point of care could improve the quality and relevance of RWD and reduce the resources needed for downstream curation and checking. Stakeholder efforts are already underway to identify and standardize key data elements that are routinely captured as part of care delivery and important for clinical research endpoint development. Multi-stakeholder initiatives such as the Minimal Common Oncology Data Elements (mCODE) and OneSource Project are pioneering new systems and approaches to improve data capture for medical product development in oncology.[6,7] Potential exists to improve data capture activities in other therapeutic areas as well, setting the groundwork and foundation for greater integration of routine care delivery and the clinical research enterprise.

The point of capture of high-quality RWD will require establishing an ecosystem of tools, standards, and workflows, and developing a shared understanding of the characteristics of high-quality data. Innovative approaches will be needed to advance the state of data collection and reduce the burden on health care providers. New tools are being developed that utilize machine learning and natural language processing, which hold promise for enabling more streamlined data capture and structured medical concept coding. New data standards are also being developed to support greater use of these tools, including the Fast Healthcare Interoperability Resources (FHIR) standard, developed by Health Level Seven (HL7), to promote data interoperability and automate the flow of clinical data sharing.

For this ecosystem to scale, however, stakeholders must value its potential and participate in efforts that advance alignment around and value for high-quality data collection. While technical solutions are needed, they are not sufficient for building the underlying culture and normative practices that will underpin these efforts. Support is needed from all levels of the health system to remove or minimize barriers, including financial or institutional obstacles, impeding the capture of high-quality data.

On July 13 and July 14, 2020, the Robert J. Margolis, MD Center for Health Policy at Duke University, in cooperation with the U.S. Food and Drug Administration, is organizing a public workshop to explore potential approaches for developing a high-quality RWD ecosystem. Discussions will consider technical challenges and cultural barriers to the scalability of data capture processes, as well as emerging practices being implemented among the stakeholder community. The following provides an overview of each session including key objectives and discussion questions the workshop will address.

[4] https://egems.academyhealth.org/articles/10.5334/egems.223/
[5] https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5983028/
[6] https://journals.sagepub.com/doi/10.1177/1740774520913819
[7] https://www.fda.gov/media/132130/download

Session 1: Identifying Data Capture Challenges and Understanding Stakeholder Needs

Notable progress has been made towards building a national data infrastructure that serves as a foundation for health care delivery. Commonly employed systems include EHRs to capture clinical information, laboratory information systems to record diagnostic and pathology data, and mobile devices that enable the sharing of patient-generated data. Leveraging these systems in pursuit of a high-quality RWD ecosystem will require more insight into current strengths and functional limitations for capturing data elements that are necessary for clinical research and evidence generation. Commonly recognized limitations include inconsistent uses of data standards, complex user design interfaces, and varying stakeholder needs that impact clinical workflow. Varying stakeholder needs may include leveraging these systems for reimbursement requirements of rendered services, ensuring confidentiality, and safeguarding provisions of protected health information. This session will overview the strengths and limitations of RWD capture tools and systems, key barriers, and current practices to improve the quality of data capture.

Discussion Questions:

- Collecting RWD involves recording information in a consistent manner using planned workflows and data capture instruments. What workflows and tools exist and how can they be characterized?
- What are the strengths of these data capture tools? What types of data are consistently captured with high reliability? What are the key limitations that prevent high-quality data capture?
- In what way is data capture burdensome and resource intensive? How could better integration of data capture and clinical workflows alleviate these burdens?

Session 2: Emerging Insights and Lessons Learned from Initiatives to Optimize Data Capture at the Point of Care

Data capture processes vary widely among health systems, which impacts the ability to leverage data in these systems for evidence needs beyond care delivery and payment. Utilizing RWD for clinical research and regulatory decision making currently requires substantial manual efforts to clean, transform, and curate RWD into an analytic dataset. Given the time and resources needed for these data curation practices, several initiatives are underway that aim to standardize core data elements routinely collected as part of clinical workflows and support data characterizations that enable and promote multiple uses of RWD elements. This session will highlight emerging initiatives such as mCODE and the OneSource Project to standardize core data sets in the oncology medical specialty with the aim of identifying key lessons learned that could inform and spur the development of the RWD ecosystem.

Discussion Questions:

- What are the critical success factors and lessons learned from establishing initiatives like mCODE and OneSource?
- What tools and workflows enable or impact core data element efforts and improve data quality?
- What is the value proposition for different stakeholders to participate in these initiatives?
- Are there key challenges that must be addressed to scale these initiatives?

Session 3: Translating Early Successes of Core Data Element Programs to Therapeutic Areas beyond Oncology

Building on the successes and progress made by initiatives in the oncology therapeutic area to improve data capture at the point of care, there is growing potential to apply similar efforts to other therapeutic areas. Fields such as ophthalmology, rheumatology, and cardiology provide examples of efforts to improve clinical workflows and data capture processes. Opening presentations will frame key considerations and opportunities to implement standardized core data sets in these therapeutic areas, and panel reactants will further explore how this standardization might apply to different data sources and elements routinely collected as part of care delivery that could support key endpoints for clinical research in diverse therapeutic areas.

Discussion Questions:

- Current initiatives in cardiology, ophthalmology, and rheumatology are developing core data elements to standardize data capture and improve quality data. What approaches are these initiatives taking that are different or similar to mCODE and OneSource?
- Would the same critical success factors identified in the previous session apply to ongoing initiatives in cardiology, ophthalmology, and rheumatology?
- What other therapeutic areas might benefit from a set of common data elements?

Session 4: Emerging Tools & Technologies to Support High-Quality Data Capture

Although the development of core sets of data elements could play an important role in defining a high-quality data ecosystem, innovation is still needed to address key gaps in workflows and -lessen the burden on providers. This session will consider both technical advances and emerging trends with data standards and tools that could further support and bolster the capture of high-quality data at the point of care. These advances and trends include machine learning and natural language processing tools, data visualization techniques, and data standards and architectures that utilize application programming interfaces (API) to improve ways in which stakeholders interact with and exchange health data. Panelists will consider how these technologies could improve existing practices and workflows, and facilitate the scalability of high-quality data capture.

Discussion Questions:

- What is the current state of adoption around emerging or novel data capture tools and data standards such as HL7 SMART on FHIR?
- What challenges might prevent the adoption or use of emerging data capture tools and limit their use across varied health systems and health care contexts?
- What is needed to scale the adoption and use of these emerging technological advancements?

Session 5: Next Steps Toward Establishing a High-Quality RWD Ecosystem

Building a high-quality RWD ecosystem will take commitment from diverse stakeholders across the health system. A shared vision is needed to align data capture tools, standards, and workflows in a way that encourages participation and returns value to stakeholders across the health system. Aligning institutional policies to support such workflows will be important along with other types of policy or financial levers that could incentivize coordinated action and continued adoption of data capture tools and processes. This session will provide key perspectives from health system leaders on what is needed

in the short, medium, and longer term as well as potential opportunities to align stakeholder value around this ecosystem.

Discussion Questions:

- What practical next steps can help establish a high-quality data ecosystem?
- What are the value propositions for establishing and participating in a high-quality data ecosystem?
- How might financing and payment mechanisms align with the need for high quality data capture?
- What other stakeholders need to be engaged to establish and promote a high-quality data ecosystem that may not have been represented at today's discussion?

Funding for this workshop was made possible in part by a cooperative agreement from the U.S. Food and Drug Administration Center for Drug Evaluation and Research. The views expressed in written workshop materials or publications and by speakers and moderators do not necessarily reflect the official policies of the Department of Health and Human Services nor does mention of trade names, commercial practices, or organizations imply endorsements by the U.S. Government.