Duke | MARGOLIS CENTER for Health Policy

Data Sharing to Accelerate Therapeutic Development for Rare Diseases

August 18 & 19, 2020

Meeting Summary

Introduction

Although rare diseases are individually uncommon they impact approximately 30 million people in the United States, or 1 in 10 Americans, outpacing the national prevalence rates of all cancers, Human Immunodeficiency Virus (HIV), and Alzheimer's Disease combined.[1] Despite the high total prevalence of rare diseases, the therapeutic development enterprise is underperforming with respect to treatment generation, with approved treatments available for only 2.4% of the existing 9,603 rare diseases.[2] The size, geographic distribution, and heterogeneous makeup of the patient population for each rare disease limit the ability of investigators to collect natural history data and sufficiently power clinical trials. These issues contribute to gaps in disease characterization, impede the conduct of robust and efficient trials, and result in sustained unmet medical need.

Existing efforts to enhance shared data resources and engage academic and industry investigators in collaborative research have shown promise in addressing gaps in disease characterization and facilitating therapeutic development. For example, researchers, funders, and regulators are building efficiencies into evidence generation by supporting the development and adoption of shared resources that improve clinical trial readiness and conduct. Researchers are also collaborating through trial networks to address priority researchable questions, increase the transparency and reproducibility of research, and leverage positive and negative learnings from clinical data collection. However, there are also several challenges associated with the use of these shared resources and collaborative networks in therapeutic development.

To discuss these challenges and potential multi-stakeholder solutions, the Duke-Margolis Center for Health Policy convened a two-day public meeting in August 2020 under a cooperative agreement with The U.S. Food & Drug Administration (FDA). Meeting participants included rare disease and therapeutic development experts from federal regulatory and research agencies, as well as experts from academic and nonprofit research institutions, patient advocacy groups, and pharmaceutical industry organizations.

The objectives of the joint meeting were two-fold—

A. To engage with industry, researchers, and other stakeholders about—
    i. the utility of shared data resources and collaborative research for addressing unmet need in therapeutic development for rare diseases,
    ii. how pharmaceutical industry investigators and other researchers are currently leveraging shared data resources and networks in preclinical research and clinical trials, and
    iii. considerations for primary data collection when the ultimate goal is to contribute data to shared resources

B. To better understand the role of FDA and other stakeholders in supporting the development, maintenance, and increased adoption of shared data resources

Meeting participants discussed approaches to enhancing and maintaining an adaptable data sharing infrastructure and to ensuring data quality, standardization, and interoperability across shared data resources. Meeting participants also discussed effective approaches for clinical trial conduct and the dissemination of research findings using collaborative networks. The following summarizes input from meeting participants regarding the primary topics of discussion at the meeting as well as potential opportunities for supporting additional progress in the use of these approaches to advance therapeutic development.

Shared Data Resources and Collaborative Research Networks as Promising Tools for Supporting Therapeutic Development in Rare Diseases

Given the impact of small population sizes on clinical trial conduct and data availability, progress in therapeutic development depends on the ability of researchers to collaborate effectively and to leverage data, tools, and findings from separate studies and shared resources. Meeting participants discussed several data sharing and collaborative research initiatives that have facilitated therapeutic discovery and development by improving understanding of disease natural history, extending the useful life of rare disease data, and enabling innovative, complementary clinical trials. For example, the IAMRARE registry program maintained by the National Organization for Rare Disorders (NORD) has supported the conduct of natural history studies that have become primary references for clinical trials and other research efforts across several rare diseases over time. The platform has supported longitudinal data collection and analysis in number of rare diseases including, Prader-Willi Syndrome and SYNGAP1-related non-syndromic intellectual disability.[3] Natural history studies developed using NORD's program have also been leveraged to recruit patients into clinical trials, to support the meaningful use of patient preference information and the development of innovative clinical outcomes assessments, and to test the performance of new clinical outcome measures.

Participants noted that collaboration through rare disease research networks supports therapeutic development because it allows researchers to compare and learn from the conduct and findings of studies implemented across institutions, therefore improving study quality and avoiding study duplication. For example, the Cystic Fibrosis Therapeutics Development Network (CFTDN) at Seattle Children's Hospital serves as a consulting hub for industry sponsored trials, providing disease-related expertise and support to ensure that trial design is congruent with regulatory standards and that data collection is standardized, comprehensive, and scientifically rigorous. Importantly, participants noted that leveraging existing data warehoused in shared resources and conducting research through collaborative networks like the CFTDN helps to minimize the burden of data collection on a limited number of rare disease patients and decreases the overall resource investment needed to support therapeutic development.

Building and Maintaining High-Quality Shared Data Resources and Collaborative Networks

To support further development and the implementation of high-quality shared data resources and productive collaborative research networks, participants noted that researchers and funders should 1) ensure that shared resources and networks are fit-for-purpose, 2) promote quality in primary data collection practices, 3) implement quality database and platform management practices, 4) engage patients and other stakeholders in resource development and study design, and 5) ensure that data and findings from new studies are contributed to shared data resources.

Ensuring Fit-For-Purpose Resources and Networks

To ensure the implementation of fit-for-purpose shared data resources and research networks, participants noted that researchers and funders should begin resource and network development with user needs and end goals in mind. For example, to achieve efficiency goals for evidence generation and to ensure the comparability of results across studies, collaborative research networks like the National Institutes of Health (NIH) Rare Disease Clinical Research Network (RDCRN) often establish centralized data management and coordinating centers (DMCCs). These DMCCs support investigators during study inception and design, data collection, and the dissemination of study results. The DMCC for the RDCRN is designed to provide data collection, administrative, and compliance tools; centralized data management capacity; and supports for study design and analyses across the network. These supports within the RDCRN have led to improved disease characterization and the regulatory approval of several therapeutics, including several treatments for urea cycle disorders.[4]

Participants discussed several other approaches for ensuring that shared data resources are fit-for-purpose. These approaches included the implementation of context-of-use agreements between database managers and data owners and users, the development of database heuristics based on discussions with end-users, and the implementation of data profiling and quality assessments to ensure data accuracy and completeness.

Additionally, participants noted that the capacity of shared data resources to ingest and integrate data from multiple, heterogenous sources was essential for ensuring that the resources were fit for the purpose of supporting efficient therapeutic development. Participants noted that the goal is for federated databases to warehouse registry data from patient groups, natural history data from academic investigators, and placebo data from industry partners, so that these data can ultimately be used together to support clinical development and reduce the need for additional, duplicative data collection efforts. Participants emphasized that data federation dictates the need for a trained workforce who can curate and maintain data warehoused in shared resources. This includes the ability of trained staff to implement quality data management standards in data ingestion and aggregation as well as the ability of staff to manage data use and transfer through relevant contracts and agreements. Finally, participants discussed the importance of approaches to assessing and ensuring the relevance of data so that it is fit-for-purpose for current research, given that data maintained in some shared data resources may be collected over a long period of time during which standard of care may have changed. Approaches to assessing the relevance of historical data for use in current clinical research include data benchmarking to standard of care and continuous assessment of the point at which longitudinal data may expire or become less informative as time passes.

Supporting Quality in Primary Data Collection

Discussion at the meeting focused extensively on the relationship between high-quality, fit-for-purpose shared data resources and high-quality primary data collection practices. Participants noted that primary data collection should be informed by study objectives and that it is critically important to implement rigorous data collection principles and practices irrespective of study characteristics. Adhering to best practices for primary data collection in natural history studies is equally as important as adhering to best practices for clinical trial data collection given that data collected across the product lifecycle can be repurposed, combined, and reused to generate evidence that supports regulatory decision making. Participants highlighted mechanisms and tools available to support the generation of regulatory-grade data including continuous discussion with FDA, through mechanisms such as Critical Path Innovation Meetings, about how natural history and other studies can be designed and implemented to facilitate clinical trial conduct and regulatory review. Participants also emphasized the benefit of standardized, comprehensive data collection practices, including the use of high-quality informed consent processes to support the conduct of innovative trials. For example, if individuals contributing their data to natural history studies or rare disease registries are consented in a way that facilitates data reuse, clinical investigators may leverage that data as part of an external or historical control group in a clinical trial, supporting efficient trial conduct and reducing data collection burden on those living with rare diseases.

Participants also discussed the importance of the harmonization of data standards across resources that support primary data collection given that researchers often face challenges retrofitting unstructured data so that it can be merged with other data to support clinical research. The adoption of common variable definitions and data collection standards, such as Clinical Data Interchange Standards Consortium (CDISC) or Fast Healthcare Interoperability Resources (FHIR) standards, supports quality in primary data collection and increases data utility by reducing the need for post-hoc standardization. Participants noted that the development and adoption of common data elements (CDEs) can also mitigate the need for extensive post-hoc data standardization and support the increased use of real-world data in therapeutic development.

Ensuring Data Fidelity, Security, and Interoperability

Meeting participants noted that data fidelity, security, and interoperability were critical dimensions for supporting the ability of shared data resources to generate tools and insights to enable efficient trial design and conduct. The implementation of high-quality database management practices is essential to ensuring data fidelity given that data from heterogenous sources with varying levels of quality are curated, stored, standardized, and aggregated within shared resources. For example, database management practices that support the collection and storage of data provenance (or data lineage) information can facilitate data transparency and composite analysis and may allow database managers to address data missingness. These practices are especially important when data is intended to be fit-for-use in the regulatory setting and compliant with requirements outlined in FDA guidance for the submission of electronic data to FDA.[5]

Providing access to patient-level data and ensuring the protection of patient privacy can be challenging in rare disease research because data anonymization is made more difficult by small population sizes. Participants noted that the implementation of best practices for data protection as well as compliance with global data security requirements (e.g. European Union Global Data Protection Regulation

requirements) are both key to supporting data security and to encouraging contributions to shared data resources. Security measures, such as techniques to guard against inadvertent re-identification and the use of independent review panels that review and approve requests for data access, mitigate the risks associated with providing access to patient-level data.[6]

Finally, participants emphasized the importance of data linkage and interoperability for improving the ability of federated databases to support improved disease characterization and efficient therapeutic development. For example, participants noted that because the likelihood of the same patient contributing to multiple data sources is high, researchers and their partners should explore and build capacity for the use of data linkage approaches, such as Global Unique Identifiers (GUIDs), to link patients with their own data across databases. These unique identifiers, along with other data linkage approaches can also be leveraged in the future to connect natural history, clinical, and longitudinal outcomes data.

Engaging Patients in Collaborative Research and the Development and Maintenance of Shared Data Resources

Understanding and addressing the goals, concerns, and needs of patients contributing their data to rare disease research is a critical component of the management of shared data resources and key to conducting clinically meaningful studies that improve therapeutic availability. Meeting participants noted that individuals who enroll in clinical trials are largely supportive of open data sharing with academia and for-profit companies. For example, in one New England Journal of Medicine study referenced at the meeting, 93% of clinical trial participants surveyed were open to sharing data with academia and 83% of those surveyed were open to sharing data with for-profit companies.[6] Additionally, the study indicated that the majority of clinical trial participants surveyed (82%) perceive that the benefits of data sharing outweigh the potential risks.[6] Participants noted that researchers should leverage this interest and meaningfully engage patient groups early to obtain input on the governance principles, goals, and primary uses of shared data resources and research networks.

Further, participants noted that meaningful patient engagement is critical to the design and conduct of patient-centered trials and to ensuring that patient input informs drug development. Participants emphasized several effective approaches to patient engagement that can facilitate study enrollment, minimize data collection burden on patients enrolled in trials, and maximize the availability of data that can be repurposed for future evidence generation. These approaches include, ensuring that patients have control of their own medical information, are consented appropriately so that their data can be reused, and are informed of trial results after completion. Finally, participants noted that continuous engagement and relationship building with patient groups is a key part of the success of research conducted through collaborative networks given that it allows for consultation throughout study design and conduct and may enable more efficient study activation in the future. For example, NIH stakeholders noted that because of existing relationships NIH's RDCRN was able to leverage an engaged network of patients and principal investigators to activate longitudinal seroprevalence surveys in rare disease patients within a matter of weeks at the onset of the COVID-19 pandemic.

Encouraging Contribution to Shared Data Resources

Participants noted that there has been a culture shift toward precompetitive collaboration and an increase in the adoption of data sharing practices as shared data resources become more widely

available and collaborative networks have shown promise in facilitating clinical trials. However, to encourage increased contributions to shared data resources, researchers, funders, and regulatory stakeholders must address a number of challenges. Researchers still face challenges associated with data standardization and interoperability that impact their ability to conduct composite data analysis and generate reproducible research. These challenges include investigator resistance to amending long-standing primary data collection practices as well as issues mapping and aligning data collection and exchange standards used in health care delivery (FHIR standards) with those used in clinical research (CDISC standards). These data alignment issues have implications for the of use of innovative trial designs and real world data in rare disease research as well as for the ability of shared data resources to support the composite data analysis that is critical to evidence generation in rare disease research. Participants also described challenges in leveraging shared data to support regulatory submissions, including challenges linked to shifts in standard of care over time, variation in the timing and conduct of safety and efficacy assessments, discrepancies and biases in trial enrollment criteria, and differences in data elements used across trials. These dimensions impact the reliability and relevance of data for supporting regulatory decision making and may limit the widespread adoption of shared data resources.

Further, to encourage and facilitate increased contribution to shared data resources, database funders and managers must accommodate patient priorities as well as the priorities of researchers from academia and industry during database and network development. Patients need to know that their data is secure and their privacy will be protected after data transfer, academic investigators need to know that their ability to publish the results of their research will be sustained, and industry stakeholders need to feel confident that data sharing is not going to place them at a competitive disadvantage. Existing data sharing approaches and mechanisms have shown promise in addressing the needs of these stakeholder groups. Effective approaches include the presence of an engaged rare disease patient community that is eager to provide input on how the data they contribute is used by the research enterprise as well as the increasing use of open access pre-print sites to facilitate the timely sharing of study results that may help to inform secondary analyses and future studies. Contributions to shared data resources have also been supported by the ability of data owners to customize and optimize their contributions in a way that protects their intellectual property and market advantage. Continuous stakeholder engagement to identify and address the needs of those who contribute to and use shared data resources will enhance their availability and capacity for supporting therapeutic development.

Key Takeaways and Next Steps

Open data sharing and effective collaborative research can support therapeutic development by allowing researchers to optimize resource use, address priority researchable questions, characterize gaps and uncertainties in scientific knowledge, and share findings and lessons learned from study conduct within and across research networks to inform future clinical trials. The optimization and increased use of these resources and networks will depend on close collaboration across patient groups, nonprofit stakeholders, academic and government researchers, and industry investigators to ensure that they are high-quality and fit-for-purpose. This collaboration should include frequent communication with FDA to facilitate continuous discussion on data quality and sufficiency throughout the therapeutic development lifecycle, including in the collection of disease natural history data. Maximizing the utility of shared data resources and networks is also contingent on taking a global approach to data collection and database governance and to addressing the priorities of a broad coalition of stakeholders around data privacy, quality, and interoperability. Accordingly, participants at the meeting outlined several next

steps for enhancing collaborative rare disease research networks and encouraging increased contribution to shared data resources, including—

a. Increasing awareness of the need for rare disease research in order to create sustained interest and continuous funding for the development of new therapeutics
b. Continuing to adopt standardized data collection processes, aligned variable definitions, and common data elements to ensure data interoperability and maximize data utility
c. Continuing to develop the capacity for the federated database management and robust analysis of large datasets
d. Continuing to involve patients and patient advocates to inform how data is collected, stored, and used to generate evidence on clinically meaningful treatment outcomes

The research community has made measurable progress in developing the infrastructure, tools, and networks critical for facilitating data sharing and collaborative therapeutic development. However, increased investment in shared therapeutic development resources, early engagement with regulatory agencies, and multi-stakeholder collaboration can increase the pace of evidence generation and improve the availability of treatments for rare diseases.

References

1. Data Sharing to Accelerate Therapeutic Development for Rare Diseases: Day 1 & 2. Margolis Center for Health Policy. Accessed March 30, 2021. https://healthpolicy.duke.edu/events/data-sharing-accelerate-therapeutic-development-rare-diseases-day-1-2

2. Ferreira CR. The burden of rare diseases. *Am J Med Genet A*. 2019;179(6):885-892. doi: https://doi.org/10.1002/ajmg.a.61124

3. Boulanger V. Overview of NORD's Registry Program. Presented at the: RDCA-DAP Launch Meeting; September 17, 2019. https://c-path.org/wp-content/uploads/2019/09/nord_registry_presentation_rdca-dap_-_vanessa_boulanger.pdf

4. Groft SC, Gopal-Srivastava R. A model for collaborative clinical research in rare diseases: experience from the Rare Disease Clinical Research Network program. *Clin Investig*. 2013;3(11):1015-1021. doi:10.4155/cli.13.101

5. U.S. Food & Drug Administration. Providing Regulatory Submissions in Electronic Format — Submissions Under Section 745A(a) of the Federal Food, Drug, and Cosmetic Act. Published online December 2014. https://www.fda.gov/regulatory-information/search-fda-guidance-documents/providing-regulatory-submissions-electronic-format-submissions-under-section-745aa-federal-food-drug

6. Institute of Medicine. Sharing Clinical Trial Data: Maximizing Benefits, Minimizing Risk. Washington, DC: *The National Academies Press.* Https://Doi.Org/10.17226/18998.; 2015. doi:10.17226/18998