Sharing Historical Trial Data to Accelerate Clinical Development

Perry T. Yin[1], Jules Desmond[1,2] and Julie Day[1,*]

Clinical trials lie at the heart of the drug development process and, although they generate vast pools of high-quality data, a large portion of this data has historically has not been well utilized. However, over the past several years, biopharmaceutical companies have entered a new era of data sharing. Herein, we describe a solution that is being developed to specifically enable data sharing across the biopharmaceutical industry and some of its early successes.

Clinical trials lie at the heart of the drug development process and, although they generate vast pools of high-quality data, historically, a large portion of this data has not been made available to other researchers. However, over the past several years, with the benefits that stem from the sharing of participant-level clinical trial data being recognized by lawmakers, biopharmaceutical companies, and researchers alike,[1,2] are marking the dawn of a new era of data sharing. Herein, we describe a solution that has been developed to enable data sharing across the biopharmaceutical industry and some of its early successes.

CREATING A GLOBAL COLLABORATIVE ECOSYSTEM

With the increase in available data sharing platforms (DSPs) within the biopharmaceutical industry, such as Clinical Study Data Request, Project Data Sphere, and Vivli, there is steady progress toward creating a future state whereby historical trial data (HTD) sharing becomes "business as usual." However, for a platform to support "business as usual" data sharing within the biopharmaceutical industry, it should be 21 Code of Federal Regulations Part 11 compliant, abide by applicable data privacy laws, and meet the highest data quality and security standards. Additionally, it must have the right governance and processes in place to facilitate sustainable growth and trust among its users.

With this in mind, TransCelerate BioPharma, a nonprofit organization currently comprised of 19 biopharmaceutical member companies (MCs), has developed a global collaborative DSP called DataCelerate. Specifically developed to support the industry's efforts to accelerate drug development through cross-company data sharing, DataCelerate was designed to be 21 Code of Federal Regulations Part 11 compliant with adherence to globally applicable data privacy and protection laws and regulations. Further, HTD being shared has undergone third-party standard formatting conversion to Standard for Exchange of Nonclinical Data (toxicology data) and Study Data Tabulation Model (uniform Clinical Data Interchange Standards Consortium (CDISC) version and a common Medical Dictionary for Regulatory Activities version for placebo/standard of care data) and is subject to quality checks upon upload through automation capabilities. The system was also designed with the needs and rigorous requirements of regulatory authorities in mind.

A CLOSER LOOK AT THE DATA BEING SHARED

DataCelerate currently houses preclinical toxicology data and, by 4Q2019, will house data from TransCelerate's Placebo and Standard of Care (PSOC) initiative. Herein, we focus on HTD being shared via the PSOC initiative.

Established in 2014, the PSOC initiative is valued as a lower-risk, high-yield opportunity to use HTD to drive efficient trial design and innovation and reduce development costs and timelines while minimizing the need to expose patients to study procedures. Participating MCs must adhere to a Data Sharing Agreement and agree to use the data only for clinical research purposes. However, unlike other DSPs, no research proposal is necessary to access the data, and data may be downloaded to and analyzed from within an MC's own internal systems. Quite purposefully, no analytic capabilities currently sit on the platform. In addition, the initiative established a blinded request process that enables data from any participating MC to be blindly requested from another MC. Each MC can approve or deny data-sharing requests; all data sharing is completely voluntary. As of June 2019, de-identified, CDISC Study Data Tabulation Model version 3.2 formatted patient-level data from 136 clinical trials involving > 85,000 patients, spanning 23 therapeutic areas is available. Additionally, over 40 different utilization examples to improve study design, conduct, and analysis have been achieved by MCs; a few examples are included herein below.

Optimizing clinical study design

Historical trial data can be used to optimize study design in multiple ways, including defining eligibility criteria, sample size, and primary and secondary end points. However, excessive or overly restrictive eligibility criteria can make trials harder to enroll, prevent otherwise eligible patients from accessing life-saving therapies, and result in trials that do not adequately represent the broader population that will be exposed to the drug. Additionally, too small a trial will result in inadequate sensitivity to detect clinical benefit, whereas an overly large

[1]TransCelerate BioPharma Inc., Conshohocken, Pennsylvania, USA; [2]Amgen Inc., Thousand Oaks, California, USA. *Correspondence: Julie Day (Julie.Day@transceleratebiopharmainc.com)

Received May 6, 2019; accepted August 7, 2019. doi:10.1002/cpt.1608

sample size leads to waste, exposing patients to unproven therapies as well as increasing development time. Last, appropriate primary and secondary end point definitions are essential in capturing the benefit of a study drug.

Following the release of updated US Food and Drug Administration (FDA) guidance defining new end points for ulcerative colitis in 2016, Eli Lilly used PSOC HTD together with real-world data to support a briefing document that was submitted to the FDA for a development program. Intense data mining was conducted to justify the proposed end points, inclusion/exclusion criteria, and sample size. By using PSOC HTD, this analysis was accomplished in 1 month, whereas the alternative, using a combination of literature, registries, and academic databases, would have taken significantly longer and added ~ $1 million in cost. Similarly, Pfizer researchers used machine-learning methods to analyze the PSOC HTD and real-world data from the Massachusetts General Hospital to address several research and development questions around early clinical drug development for intracerebral hemorrhage.[3,4] These included the linkage between the mechanistic end point and clinical outcomes, the identification and confirmation of the predictors of hematoma expansion, functional outcomes, and mortality in intracerebral hemorrhage. This information was used to inform study design, identify the right treatment population, assess study feasibility, and ultimately expedite the clinical trial decision process.

Informing clinical analysis

Late-stage clinical trials can fail or be put on hold due to product safety-related questions that arise during the safety review process. To better evaluate safety events, an understanding of the type and background incidence of adverse events (AEs) specific for the population being studied is required, allowing for proper characterization of the event of interest. However, the large observational databases from which AE incidence has traditionally been derived have limits.

Addressing this, researchers are now combining traditional databases with HTD where AE data were rigorously collected, providing valuable contextual safety information that not only leads to better patient protection but also has the potential to save drug development costs. Researchers at Genentech recently used PSOC HTD from rheumatoid arthritis trials to assess potential regional differences in the rates of American College of Rheumatology (ACR) Criteria response and in AE reporting.[5] Data from seven trials in the PSOC database were analyzed, wherein patients were grouped by region (i.e., Asia, Latin America, and the Russian Federation and Eastern Europe) and evaluated for differences in demographics, AE reporting rates, and ACR response. This revealed significant regional differences in AE reporting rates and ACR50/ACR20 response rates, with differences in Latin America, Russian Federation and Eastern Europe, and Asia being especially notable. As such, it was concluded that future patient populations from these regions may show distinct efficacy/safety profiles regardless of treatment and that managing recruitment by region to balance these factors may be warranted to prevent biases and incorrect inferences from being drawn around efficacy and/or safety.

THE FUTURE IS DATA SHARING

We envision a future where data sharing is business as usual and fundamental to how clinical research is designed and conducted. To achieve this future state, it will take a collaborative, multipronged approach involving regulatory authorities, researchers, and biopharmaceutical companies to address current challenges in fully utilizing HTD (e.g., minimizing selection bias and ensuring traceability) and making HTD sharing the norm. In an effort to maximize the value of HTD, stimulate novel research, and help accelerate the development of medicines for patients, TransCelerate is working to address these challenges while looking ahead to determine what other use cases, data types, and/or connectivity among both existing data types (i.e., linking data) and/or other DSPs may bring additional value to DataCelerate and the clinical research ecosystem it aims to support.

ACKNOWLEDGMENTS

The authors would like to acknowledge Xiang Zhang and the Eli Lilly team for providing the ulcerative colitis example, Florence Yong and the Clinical Drug Development Team led by Steven Arkin at Pfizer, and the Massachusetts General Hospital study team led by Joshua Goldstein for contributing the intracerebral hemorrhage example, and Alun Bedding and the Roche/Genentech team for providing the rheumatoid arthritis example. We would also like to thank the members of the Placebo and Standard of Care Data Sharing Initiative at TransCelerate Biopharma, including Ed Bowen, Glenn Herbert, Jessica Lim, Aaron Mann, Ulo Palm, Virginia Nido, and Saba Tahseen, as well as David Bobbitt, Robert Beckman, Kathryn Hall, and Mary Drake Dixon for their insights and suggestions.

FUNDING

No special funding was provided in the authoring of this paper; research and authoring was funded through normal TransCelerate dues.

CONFLICT OF INTEREST

Perry Yin was a consultant contracted by TransCelerate Biopharma. Julie Day is an employee of TransCelerate Biopharma. Jules Desmond is a member of TransCelerate Biopharma.

1. Ross, J.S. & Krumholz, H.M. Ushering in a new era of open science through data sharing: the wall must come down. *JAMA* **309**, 1355–1356 (2013).
2. Warren, E. Strengthening research through data sharing. *N. Engl. J. Med.* **375**, 401–403 (2016).
3. Cai, X. *et al*. Predictive modeling of outcomes in intracerebral hemorrhage for clinical trial design. Neurocritical Care Society 16th Annual Meeting, Boca Raton, FL, September 28, 2018.
4. Yong, F.H., Li, R., Hua, S.Y., Palmer, J. & Bugarini, R. Utilization of historical data and real world evidence in clinical trial development – case studies in rare disease and oncology. Joint Statistical Meeting, Vancouver, Canada, July 28–August 2, 2018.
5. Keebler, D. *et al*. Adverse event reporting rates and placebo standard-of-care-arm ACR responses vary by region in rheumatoid arthritis trials. *Ann. Rheum. Dis.* **77**, 74 (2018).