



PhUSE US Connect 2019

Paper AB06

The Power of Data Sharing: Real Projects with Real Impacts

Dave Handelsman, Project Data Sphere, LLC, Cary, NC, USA

ABSTRACT

Data sharing leading to improved cancer patient outcomes provides exciting examples of applying Big Data and Analytics to clinical research. Since 2014, Project Data Sphere has aggregated patient-level data from more than 135,000 patients spanning 150 trials, and made this data available for exploration and investigation to more than 2,000 researchers.

This successful data sharing program has enabled medically relevant insights to be discovered and published in leading peer-accepted journals, and new research programs are underway, including:

- Development of an external control arm for small cell lung cancer that may eliminate the need to enroll cancer patients in standard-of-care comparator treatment arms
- Development of machine learning tools for tumor image processing
- Investigations into rare but serious adverse events associated with new of immuno-oncology therapies

In this session, the opportunities and challenges of clinical trial data sharing will be discussed as they relate to Big Data and Analytics.

INTRODUCTION

Data sharing has broad meaning, even when discussed in the context of clinical research. Depending on the context, this concept can refer to the sharing of patient-level data traditionally associated with clinical trials, or it can alternatively (or additionally) refer to less traditional data types such as genomics and images, and can also refer to data extracted from electronic health records.

For the purposes of this paper, data sharing will be discussed in the context of data made available from completed clinical trials. While there are specific technical discussions that are best addressed in a different paper, the focus, here, instead, will be on how the data sharing community, which includes biopharmaceutical companies, academic and government research organizations, researchers, and, ultimately, patients, have overcome the initial hurdles associated with data sharing, and progressed to the point where impactful new science is being derived via data sharing initiatives.

A LITTLE HISTORY

Data sharing is not a new concept. For many years, interested researchers have been able to request patient-level datasets from clinical trial sponsors on an as-needed basis to support their research interests. During this time, there was not a truly common process in place and adopted by biopharmaceutical companies, and each request was typically assessed based upon the merits of the research proposal, the qualifications of the investigator(s) and status of the data being requested. Once a request was approved, the data was provided to the researchers, typically for a specified duration, during which time the investigation was expected to be completed.

In 2013, however, there was emerging interest in developing and implementing a grander approach to data sharing. Largely stemming from policy plans of the *European Medicines Agency*¹, and then backstopped by the *EFPIA-PhRMA Principles for Responsible Clinical Trial Data Sharing*², various global biopharmaceutical companies embarked on a series of data sharing programs. In some cases, a common data sharing platform, www.clinicalstudydatarequest.com, that spanned multiple biopharmaceutical companies was implemented, and in other cases individual biopharmaceutical companies deployed their own data sharing platforms. Additional independent data sharing environments that were developed include the *Project Data Sphere* cancer research platform (www.projectdatasphere.org), the *NCTN/NCORP Data Archive* (<https://nctn-data-archive.nci.nih.gov/>) and, most recently, *Vivli* (www.vivli.org).



Each of these platforms varies in their implementation approach, but they were largely faced with tackling the same technical, legal, process and cultural hurdles – not the least of which is the historic resistance to broadly sharing clinical data. This resistance was highlighted in the *New England Journal of Medicine* editorial³ from January 2016 which stated that “there is concern among some front-line researchers that the system will be taken over by what some researchers have characterized as ‘research parasites’”.

The current availability of multiple data sharing platforms, and population of these platforms with true patient-level clinical trial data that spans indications and organizations, demonstrates that much of the angst present at the onset of data sharing programs has dissipated in just a few short years. Most mature biopharmaceutical companies have established central data sharing offices, with formalized data sharing processes, and less mature biopharmaceutical companies are evaluating how best to join the data sharing community. In many ways, the core hurdles have been overcome, but there is still work to be done to fully leverage the value of shared clinical trial data.

IF YOU BUILD IT, WILL THEY COME?

For the first several years of these formalized data sharing programs, there had been growing concern that the aggregation of patient-level clinical trial data had been a time-consuming expense without recognizable value. In some ways, this is an accurate assessment. The burden of preparing datasets to be shared falls on biopharmaceutical companies with regard to the time and effort required to de-identify (or anonymize) the data, provide any necessary data clean-up and curation, obtain the necessary legal and business approvals, etc. That work, however, does not necessarily yield a direct benefit for the data provider. It is primarily, or even exclusively, an expense, and one that frequently distracts from other business priorities. The implementation of dedicated data sharing teams within these companies, however, points to the recognition that biopharmaceutical companies appreciate data sharing as a necessary business expense.

With data sharing still in its early years, and most formalized data sharing initiatives occurring after the 2013/2014 EMA and EFPIA/PhRMA activities, there have been several years where the primary activity with regard to data sharing has focused on process development, and ultimately, data provisioning. In that timespan, new scientific insights derived from the various data sharing programs have been relatively few. This is, not surprisingly, a function of both the relative sparsity of the data since it takes time to develop the necessary processes and then to prepare the data for sharing, and, in general, relatively low-level awareness among potential researchers.

In short, the best mechanism for attracting researchers was hobbled by two critical factors. The data sharing platforms were in the earliest stages of maturity with regard to data content, and the publication of new insights which would create excitement and interest among potential researchers was even less mature.

SUCCESS BREEDS SUCCESS

In 2015, *Project Data Sphere, LLC* launched a public-facing research program – the *Prostate Cancer DREAM Challenge*⁴. This research program used curated data from four phase III prostate cancer trials available through the *Project Data Sphere* cancer research platform, and invited crowdsourcing participants to predict overall survival for these patients using the patients’ baseline clinical variables, and to predict treatment discontinuation due to adverse events.

While the *DREAM* platform had been used to host multiple crowd-sourcing challenges previously, the *Prostate Cancer DREAM Challenge* attracted more *DREAM* solvers to date than any previous challenge. The best performing solvers developed exciting improvements to existing prognostic models. These improved prognostic models have the potential to directly impact patient care – enabling physicians to use a patients’ clinical measurements to more accurately predict their outcomes and adjusting their treatment plan accordingly, and exploring whether patient engagement with regard to potential (or actual) adverse events can be used to encourage patients to continue their prescribed course of treatment.

The success of the *Prostate Cancer DREAM Challenge*, however, is not limited to the new insights regarding patient outcomes, and their potential impact on patient care. These new insights, accompanied by peer-accepted publications and the associated publicity, spurred new interest in data sharing for both data providers and researchers. Data providers gained confidence that their efforts to prepare and deliver data for data sharing initiatives can have demonstrable impacts in the patient community, and additional researchers became aware of valuable new data sources to drive their research interests. Importantly, the successes associated with the *Prostate Cancer DREAM Challenge* additionally have led to new funding opportunities, which, in turn, are leading to new research programs.



THE BEST IS YET TO COME

In its brief history, data sharing was initially driven by the “if you build it, they will come” approach, and there has been some demonstrated success through this strategy as discussed previously. However, learnings over the past several years have shown that targeted research can be a much better driver of data attainment and aggregation, and subsequent new science and insights. Rather than soliciting “data”, in general, from biopharmaceutical companies, a more robust strategy is to develop targeted data demands associated with specific research programs.

In the case of data sharing efforts like *ClinicalStudyDataRequest.com*, this can mean that although vast amounts of clinical trial data are pledged to the repository, approved research requests initiated by investigators drive the availability of that data within the data sharing platform to support the research request goals. In this case, an independent review board assesses the merits of the research proposal, and renders a decision indicating their approval or rejection of the proposal. From that point forward, data associated with approved research projects are recruited to the data sharing platform. A summary of published research proposals associated with *ClinicalStudyDataRequest.com* can be found at <https://clinicalstudydatarequest.com/Metrics/Published-Proposals.aspx>.

Project Data Sphere, LLC has seen the success of the *Prostate Cancer DREAM Challenge* lead to a series of peer-accepted publications that are a direct result of the *DREAM* challenge, and the success of that research program has inspired additional independent research and subsequent peer-accepted publications regarding pancreatic, non-small cell lung cancer and colorectal cancer, among other research topics. A complete list of peer-accepted publications can be found in the [Appendix](#).

The recognized success of the *Prostate Cancer DREAM Challenge* has recently spurred the launch for four *Project Data Sphere*-managed data-driven research programs. These programs, briefly summarized here, are at various stages of maturity, and are designed to not only drive new insights regarding cancer patient outcomes, but to also spur the delivery of new data to the *Project Data Sphere* cancer research platform. These new data are expected to not only power the research program to which they are directed, but to also add to the overall repository of data to drive new, currently unplanned, research. In effect, Project Data Sphere, LLC is implementing a virtuous engine where new research programs drive new data aggregation, which drives new research, and so on.

EXTERNAL CONTROL ARM RESEARCH PROGRAM

In cancer trials, patients typically receive either the experimental treatment or a standard-of-care treatment. This research program, which will initially begin with small cell lung cancer, is focused on using previously collected clinical trial data to develop a control arm to be used in place of treating patients with a standard-of-care therapy as part of a new clinical trial. A successful external control arm will enable fewer patients to be enrolled in trials (reducing cost and time), will enable all trial participants to receive the experimental treatment, and will enable those patients that would have received standard-of-care treatment based on traditional trial designs to potentially receive experimental treatment in other trials (thereby reducing the cost and time of those trials).

IMAGING AND ANALYTICS

Image analysis during cancer clinical trials is a costly and time-consuming process, frequently reliant on image reads by multiple reviewers to arrive at a consensus assessment. This research program is focused on the development of machine-learning algorithms that will enhance the performance of imaging in clinical trials through greater reliability, automation and outcomes of response assessment metrics. As with the *Prostate Cancer DREAM Challenge*, this research program will be organized as a crowdsourcing challenge open to the public.

RARE TUMOR REGISTRIES

Prospective tumor registries have the potential to complement conventional clinical trials which, through randomized controlled design, attempt to minimize threats to internal validity. Due to the strict and often narrow eligibility criteria of clinical trials, however, they often have deficits in external validity. Through purposeful prospective capture of real-world data at point-of-care, the rare tumor registry program, has the potential to obtain a much broader representation of rare patient cases. This broader representation is essential to advance the development of drug therapies for rare but deadly tumor types. The real-world data captured during this long-term registry program, which will initially be focused on Merkel Cell Carcinoma, will be periodically delivered to the *Project Data Sphere* cancer research platform to support broad research where the data is expected to be investigated in conjunction with data from rare tumor clinical trials.



IMMUNE-RELATED ADVERSE EVENTS

Immunotherapy is a cancer treatment strategy that harnesses the innate power of the body's immune system to recognize and attack cancer cells that have been hiding and target them for destruction⁵. As of today, the FDA has approved seven immune checkpoint inhibitors (ICIs) since the introduction of ipilimumab in 2011 to treat advanced melanoma, and in many cases the patient outcomes for this new class of treatment have been remarkable. Although all treatments carry risks, there is growing recognition that ICIs have the additional potential for serious but rare immune-related adverse events (irAEs). This research program is focused on identifying these irAEs, developing valuable insights into the management of irAEs early in drug development, potentially facilitating clinical trials and decreasing the chance that promising treatments are limited by toxicity.

CONCLUSION

Extraordinary progress regarding patient-level clinical trial data sharing has occurred since the initial first steps were taken in 2013, just over 5 years ago. Data sharing has evolved, conceptually from "Why?" to "When?", and impactful data science that can improve patient care is being driven via these data sharing programs while they continue to grow in data content, in platform capabilities, and in research-driven outcomes. In parallel, data sharing efforts regarding real-world data, genomics and imaging continue to progress and contribute the ecosystem of patient-level data available for new research. Each of these data types provides its own challenges and opportunities as we collectively work toward a comprehensive data sharing community.

REFERENCES

1. *European Medicines Agency policy on publication of clinical data for medicinal products for human use*, https://www.ema.europa.eu/documents/other/european-medicines-agency-policy-publication-clinical-data-medicinal-products-human-use_en.pdf, October 2014.
2. *EFPIA-PhRMA Principles for Responsible Clinical Trial Data Sharing*, <http://phrma-docs.phrma.org/sites/default/files/pdf/PhRMAPrinciplesForResponsibleClinicalTrialDataSharing.pdf>, 2013.
3. *Data Sharing*, Dan L. Longo, MD, and Jeffrey M. Drazen, MD, <https://www.nejm.org/doi/full/10.1056/NEJMe1516564>, January 2016.
4. The Prostate Cancer DREAM Challenge, <https://www.synapse.org/#!/Synapse:syn2813558/wiki/70844>.
5. www.PatientResource.com.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the author at:

Dave Handelsman
Project Data Sphere, LLC
Cary, NC
919 531 2694
Dave.Handelsman@ProjectDataSphere.org
www.linkedin.com/in/davehandelsman
<https://twitter.com/dhandelsman>
www.ProjectDataSphere.org

Brand and product names are trademarks of their respective companies.

APPENDIX

	Title	Author	Publication / Publication Date
1	Statin Treatment and Outcomes of Metastatic Pancreatic Cancer: A Pooled Analysis of Two Phase III Studies	Omer Abdel-Rahman	Clinical and Translational Oncology November 2018
2	5-Fluorouracil-Related Cardiotoxicity: Findings from Five Randomized Studies of 5-Fluorouracil-based Regimens in Metastatic Colorectal Cancer	Omer Abdel-Rahman	Clinical Colorectal Cancer November 2018
3	Obstruction Predicts Worse Long-term Outcomes in Stage III Colon Cancer: A Secondary Analysis of the N0147 Trial	Fadi S. Dahdaleh, Scott K. Sherman, Elizabeth C. Poli, Janani Vigneswaran, Blase N. Polite, Manish R. Sharma, Daniel V. Catenacci, Steven B. Maron, Kiran K. Turaga	Surgery October 2018
4	Data Integration Innovations to Enhance Analytic Utility of Clinical Trial Content to Inform Health Disparities Research	Stephen B. Cohen, Jennifer Unangst	Frontiers in Oncology September 2018
5	ePCR: An R-package for Survival Time and Time-to-Event Prediction in Advanced Prostate Cancer, Applied to Real-World Patient Cohorts	Teemu D. Laajala, Mika Murtojärvi, Arho Virkki, Tero Aittokallio	Bioinformatics June 2018
6	Prediction of Overall Survival for Metastatic Pancreatic Cancer: Development and Validation of a Prognostic Nomogram with Data from Open Clinical Trial and Real-World Study	Junjie Hang, Lixia Wu, Lina Zhu, Khigiang Sun, Ge Wang, Jingjing Pan, Suhua Zheng, Kequn Xu, Jiadi Du, Hua Jiang	Cancer Medicine June 2018
7	Myocarditis Associated with Immune Checkpoint Inhibitors: An Expert Consensus on Data Gaps and a Call to Action	Tomas G. Neilan, Mace L. Rothenberg, Laleh Amiri-Kordestani, Ryan Sullivan, Richard M. Steingart, William Gregory, Subramanian Hariharan, Tarek A. Hammad, Joann Lindenfeld, Martin J. Murphy, Javid Moslehi, On Behalf of the Checkpoint Inhibitor Safety Working Group	The Oncologist 25 May 2018
8	On the Relationship Between Tumour Growth Rate and Survival in Non-Small Cell Lung Cancer	Hitesh B. Mistry	PeerJ November 2017
9	Effect of Breast Irradiation on Cardiac Disease in Women Enrolled in BCIRG-001 at 10-Year Follow-Up	S. Peter Wu, Moses Tam, Raymond Mailhot Vega, Carmen A. Perez, Naamit Kurshan Gerber	Radiation Oncology November 2017
10	A DREAM Challenge to Build Prediction Models for Short-Term Discontinuation of Docetaxel in Metastatic Castration-Resistant Prostate Cancer	Fatemeh Seyednasrollah, Devin C. Koestler, Tao Wang, Stephen R. Piccolo, Roberto Vega, Russell Greiner, Christiane Fuchs, Eyal Gofer, Luke Kumar, Russell D. Wolfinger, Kimberly Kanigel Winner, Chris Bare, Elias Chaibub Neto, Thomas Yu, Liji Shen, Kald Abdallah, Thea Norman, Gustavo Stolovitzky, Howard R. Soule, Christopher J. Sweeney, Charles J. Ryan, Howard I. Scher, Oliver Sartor, Laura L. Elo, Fang Liz Zhou, Justin Guinney, James C. Costello, and Prostate Cancer DREAM Challenge Community	JCO Clinical Cancer Informatics August 2017

	Title	Author	Publication / Publication Date
11	Sounding Board: Advantages of a Truly Open-Access Data-Sharing Model	Monica M. Bertagnolli, Oliver Sartor, Bruce A. Chabner, Mace L. Rothenberg, Sean Khozin, Charles Hugh-Jones, David M. Reese, Martin J. Murphy	New England Journal of Medicine March 2017
12	Assessment of a Prognostic Model, PSA Metrics and Toxicities in Metastatic Castrate Resistant Prostate Cancer using Data from Project Data Sphere.	Bethany Pitcher, Leila Khoja, Robert J. Hamilton, Kald Abdallah, Melania Pintilie, Anthony M. Joshua	PLOS One February 2017
13	Estimation of Tumour Regression and Growth Rates During Treatment in Patients with Advanced Prostate Cancer: A Retrospective Analysis.	Julia Wilkerson, Kald Abdallah, Charles Hugh-Jones, Greg Curt, Mace Rothenberg, Ronit Simantov, Martin Murphy, Joseph Morrell, Joel Beetsch, Daniel J Sargent, Howard I Scher, Peter Lebowitz, Richard Simon, Wilfred D Stein, Susan E Bates, Tito Fojo	Lancet Oncology December 2016
14	Prediction of Overall Survival for Patients with Metastatic Castration-Resistant Prostate Cancer: Development of a Prognostic Model Through a Crowdsourced Challenge with Open Clinical Trial Data.	Justin Guinney, Tao Wang, Teemu D Laajala, Kimberly Kanigel Winner, J Christopher Bare, Elias Chaibub Neto, Suleiman A Khan, Gopal Peddinti, Antti Airola, Tapio Pahikkala, Tuomas Mirtti, Thomas Yu, Brian M Bot, Liji Shen, Kald Abdallah, Thea Norman, Stephen Friend, Gustavo Stolovitzky, Howard Soule, Christopher J Sweeney, Charles J Ryan, Howard I Scher, Oliver Sartor, Yang Xie, Tero Aittokallio, Fang Liz Zhou, Dr James C Costello	Lancet Oncology November 2016
15	“Threshold-crossing”: A Useful Way to Establish the Counterfactual in Clinical Trials?	H-G Eichler, B Bloechl-Daum, P Bauer, F Bretz, J Brown, LV Hampson, P Honig, M Krams, H Leufkens, R Lim, MM Lumpkin, MJ Murphy, F Pignatti, M Posch, S Schneeweiss, M Trusheim, F Koenig	Clinical Pharmacology & Therapeutics October 2016
16	Predicting Survival of Pancreatic Cancer Patients Treated with Gemcitabine Using Longitudinal Tumour Size Data.	Thierry Wendling, Hitesh Mistry, Kayode Ogungbenro, Leon Aarons	Cancer Chemotherapy and Pharmacology May 2016
17	A Patient-Level Data Meta-Analysis of Standard-of-Care Treatments from Eight Prostate Cancer Clinical Trials.	N. Geifman, A.J Butte	Nature Scientific Data May 2016
18	Individual Patient Data Analysis of Randomized Clinical Trials: Impact of Black Race on Castration-resistant Prostate Cancer Outcomes.	Daniel E. Spratt, Yu-Wei Chen, Brandon A. Mahal, Joseph R. Osborne, Shuang G. Zhao, Todd M. Morgan, Ganesh Palapattu, Felix Y. Feng, Paul L. Nguyen	European Urology April 2016
19	Comparative Effectiveness of Mitoxantrone plus Prednisone versus Prednisone alone in Metastatic Castrate-resistant Prostate Cancer after Docetaxel Failure.	Angela K. Green, Robert W. Corty, William A. Wood, Mathew Meeneghan, Katherine E. Reeder-Hayes, Ethan Basch, Matthew I Milowsky, Stacie B. Dusetzina	The Oncologist May 2015