Contemporary Clinical Trials Communications 17 (2020) 100515

Contents lists available at ScienceDirect

Contemporary Clinical Trials Communications

journal homepage: http://www.elsevier.com/locate/conctc

Opinion paper

Considerations for using distributed research networks to conduct aspects of randomized trials

Keith A. Marsolo [a,*], Jeffrey S. Brown [b], Adrian F. Hernandez [c], Bradley G. Hammill [d], Sudha R. Raman [a], Beth Syat [e], Richard Platt [b], Lesley H. Curtis [d]

[a] Department of Population Health Sciences, Duke University, Durham, NC, 27701, USA
[b] Department of Population Medicine, Harvard Medical School, Boston, MA, 02215, USA
[c] Department of Medicine, Duke University School of Medicine, Durham, NC, 27710, USA
[d] Department of Population Health Sciences, Department of Medicine, Duke University, Durham, NC, 27701, USA
[e] Department of Population Health Medicine, Harvard Medical School, Boston, MA, 02215, USA

A B S T R A C T

Stakeholders in the clinical research enterprise are aligned around the need to make clinical research in general, and randomized controlled trials in particular, more meaningful and efficient. To that end, we have built distributed research networks (DRNs) for the Sentinel System, the National Institutes of Health (NIH) Collaboratory, and the National Patient-Centered Clinical Research Network (PCORnet). DRNs reuse electronic health record (EHR) and claims data for research. The design and use of health data DRNs is complicated by lack of uniformity in data collection, a fragmented healthcare system, and the imperative to protect research participants.

We describe the key elements of successful DRNs, as well as methods, challenges, and solutions we have encountered in using DRNs to support different phases of randomized, multi-site, clinical research. This work supports "real-world" efforts to build a learning health system and will enable others to conduct randomized clinical trial research using a DRN.

1. Introduction

Distributed research networks (DRNs) support the conduct of large-scale patient-centered outcomes research and comparative effectiveness research, including the Sentinel Initiative's Food and Drug Administration (FDA)-Catalyst program, the National Patient-Centered Clinical Research Network (PCORnet®) and the National Institutes of Health (NIH) Health Care Systems Collaboratory (Collaboratory; Table 1). These networks use data from insurance claims and electronic health record sources to support studies of pharmacovigilance and medication effectiveness, long-term observational follow-up, and pragmatic clinical trials.

The 21st Century Cures Act, passed in the United States in 2016, pushed regulators to define guidelines so that real-world data could serve as the source of real-world evidence, thereby accelerating existing efforts to leverage real-world data for decion-making. Pursuant to this Act, in December 2018, FDA published a framework for using real-world data for evidence generation in order to promote shared learning and guide drug development [2]. To derive real-world evidence from real-world data, information available throughout the healthcare system—in the electronic health record and insurance claims—is repurposed for other uses.

Every DRN is different in its focus and purpose, but they all share some commonalities, including harmonization of data into a Common Data Model (CDM) and a process of curation to ensure data are fit for purpose before being used for research (see Table 2 for a summary of key components). Through these activities, DRNs have gained tremendous knowledge and experience on the complexities real-world data. Two of the biggest challenges with real-world data include a lack of uniformity collection within and across healthcare systems, along with fragmented care delivery that often means that no single party has a complete view of all of the variables or outcomes of interest for the entire study. DRNs have developed methods to identify and mitigate these issues, and these mitigation costs can be spread across multiple studies, making them a more cost-effective platform than starting from scratch as part of a single study.

We, along with our colleagues, have built DRNs for the FDA-Catalyst program, the NIH Collaboratory, and PCORnet in close alignment with

* Corresponding author. 200 Morris St, DUMC 3850, Location 7424, Durham, NC, 27701, USA.
E-mail addresses: Keith.Marsolo@duke.edu (K.A. Marsolo), jeff_brown@harvardpilgrim.org (J.S. Brown), adrian.hernandez@duke.edu (A.F. Hernandez), brad.hammill@duke.edu (B.G. Hammill), sudha.raman@duke.edu (S.R. Raman), beth_syat@harvardpilgrim.org (B. Syat), richard_platt@harvard.edu (R. Platt), lesley.curtis@duke.edu (L.H. Curtis).

https://doi.org/10.1016/j.conctc.2019.100515
Received 13 August 2019; Received in revised form 18 December 2019; Accepted 30 December 2019
Available online 2 January 2020

Table 1
Distributed research networks.

DRN (initiation date)	Sentinel (2009)	Collaboratory (2014)	PCORnet (2014)
Stated Goal	Assess the use, safety, and effectiveness of regulated medical products by using EHR data and other resources.	To facilitate multi-site research collaborations between investigators and data partners by creating secure networking capabilities and analysis tools for electronic health data	To improve the nation's capacity to conduct rapid, efficient, and economical comparative effectiveness research.
Data Source [1]	Administrative claims and demographic data, outpatient pharmacy dispensing data, and some lab tests and vital signs data.	The Sentinel DRN provided the foundation, supplemented by EHR and claims data (as needed for longitudinal data capture), and patient-reported outcomes	EHR data supplemented by claims data (as needed for longitudinal data capture), and patient-reported outcomes (PROs)
N (as of December 2019)	~310 million unique member IDs. 668 million years of observation time. 70 million people accruing new data.	~45 million available for clinical trials	~37 million available for clinical trials
Data elements	Medical encounters: ambulatory care diagnoses and procedures, outpatient pharmacy dispensing, laboratory testing (procedure order) and selected test results, inpatient diagnoses, treatments and procedures itemized on a hospital bill	Similar to Sentinel	Similar to Sentinel, but with data sourced from EHRs instead of administrative claims sources. Also includes inpatient and outpatient medication administrations, medication prescribing, laboratory results, vital statistics, PROs and clinical and general observations.
Coordinating Centers*	Harvard Pilgrim Health Care Institute	Duke Clinical Research Institute Harvard Pilgrim Health Care Institute Kaiser Permanente Washington Health Research Institute Johns Hopkins Berhman Institute of Bioethics	Duke Clinical Research Institute Harvard Pilgrim Health Care Institute National Patient-Centered Clinical Research Network (PCORnet) Genetic Alliance (2014–2019)
Initial funding source	Created in response to a congressional mandate in the FDA Amendments Act (FDAAA) of 2007	The Common Fund at the NIH	The Patient-Centered Outcomes Research Institute (PCORI).
Number of network partners	35	15	11

Table 2
Key elements of a DRN.

Element	Description
Governance	Within a DRN, governance will cover the structure of the network, how network partners interact with the Coordinating Center, expectations around data and query response, policies around authorship, privacy, security, patient engagement, etc. In addition, the governance structure will dictate how sponsors and investigators interact with the DRN and the network partners, and in the case of a randomized trial, the patients who will be recruited and enrolled [3].
Common Data Model (CDM)	Partners transform and harmonize their electronic data into a common data format.
Data Curation	For most DRNs, data are checked for conformance to the data model and against quality metrics related to plausibility, completeness and persistence at both the partner/site and at the coordinating center to determine if they are ready for analysis. Other checks are augmented with focused assessment of data elements and concepts that are critical to each study.
Secure Querying Platform and Architecure	DRNs employ software to securely distribute queries and receive responses. The three DRNs described above use PopMedNet. DRNs that opt for an asynichronous querying approach allow network partners to maintain physical and operational control of their data, enabling sites to share only the minimum necessary data for each study. DRNs that allow synchronous querying and response may give network partners less control over the uses of their data and the information included in responses.

academic research organisations and participating health care systems (network partners). In this article, we describe how distributed research networks can facilitate different phases of a research study, specifically randomized clinical research.

2. Leveraging a DRN for a randomized clinical trial

DRNs provide conduits to underlying populations, which will vary depending on whether the network is based on a health plan (i.e., FDA Catalyst and Collaboratory) or a health system (i.e., PCORnet). Therefore, the kinds of questions an investigator can ask, and subsequently the kind of trial that can be done, will differ based on the type of DRN. Below, we describe considerations for using a DRN to support different aspects of the design and conduct of a randomized clinical trial.

2.1. Feasibility assessment

An investigator can use a DRN to answer specific questions related to study feasibility, such as *What are the observed patterns and distribution of a condition or intervention? What is the rate of an outcome during exposure to a drug? How can treatment outcomes be measured? What is the potential size of the trial-eligible population and what are the observed outcomes rates?* Examining patterns of real-world utilization, medication use, and clinical care for a specific population can help an investigator with site identification by determining whether a network partner has the data required for the trial, as well as in tailoring study inclusion and exclusion criteria so they reflect the way data are captured in the real-world, which eases the process for participant screening. Going further, feasibility assessments can include comparisons across partners/sites, which can help uncover data quality issues, further informing decisions about site identification. Examples of such assessments include: *For a given population, what are event rates across sites for procedures, medication orders, outcomes? Are differences in rates due to practice variation or missing/incorrectly mapped data?*

2.2. Clinician engagement

Studies that will test a randomized intervention involving a prescription or other therapy requiring a provider will likely need to partner with a DRN with access to health system data as well as the clinicians who practice in those systems. These types of intervention studies require the affirmative engagement of the network partners and standardized approaches for randomization and intervention. Typically, they will require that the topic being studied be of interest to the system or clinical stakeholders, that the outcome have the potential to improve the care they provide, that there be an internal champion at each site, that the trial not interfere unduly with the routine delivery of care, and that the trial not carry reputational risk for the organization [4,5].

2.3. Patient identification/recruitment

As part of the feasibility assessment, computable phenotypes may be deployed in order to identify patients of interest based on current health status or medical history. Depending on internal recruitment workflows at participating sites, it may also be necessary to implement these phenotypes within the operational systems, for instance, to alert a provider that the patient they are seeing in clinic that day is potentially eligible for the trial, or to contact them through the health system's patient portal.

A DRN based on health plan data cannot recruit participants through providers, but can still recruit people by email, mail, and phone. This is not always ideal, as a participant may be reluctant to participate in research if the invitation comes from their insurance provider instead of their healthcare provider. For some interventions, such as educational interventions, using a health plan DRN is a reasonable approach. For example, the vanguard randomized study from the FDA Catalyst program is the IMplementation of an RCT to imProve Treatment With Oral AntiCoagulanTs in Patients With Atrial Fibrillation (IMPACT-AFib) trial, which is designed to improve oral anticoagulant therapy in patients with atrial fibrillation [6]. Oral anticoagulation therapy reduces risk of stroke in patients with atrial fibrillation, and yet only about half of these patients take this medication [6]. The intervention is a direct-mail educational intervention that is distributed through the health plan. The study team will also contact the providers to let them know that material was sent to the patients and to educate the providers [6].

The patients are randomized to either receive the educational intervention or standard of care. Because randomization to receive the mailer does not require a prescription, and consequently a provider, randomization can occur through the health plan DRN. The Aspirin Dosing: A Patient-Centric Trial Assessing Benefits and Long-term Effectiveness (ADAPTABLE) trial, a large, pragmatic trial designed to determine the optimal dose of maintenance aspirin for patients with coronary artery disease, (N = ~15,000 patients; 20) is another study that is using health plan-based recruitment, though the majority of patients were recruited through health systems [7,8].

2.4. Enrollment

Investigators must develop processes to enroll patients in trials and to notify the DRN. As with any trial, activated and engaged clinicians and health system leaders are critical for achieving enrollment goals. Internet-based tools, such as online portals for enrollment, can substantially reduce the overhead of conducting multi-site trials. ADAPTABLE uses the PCORnet DRN and has an online portal for consent and randomization that can also be used for patient follow-up. One way that participants can be connected to the DRN is through a crosswalk table in the CDM that contains a link between the trial identifier and local datamart identifiers for enrolled patients. This enables periodic extracts of electronic health record or claims data for the study teams.

2.5. Assessing complete outcomes

Within a trial, an investigator needs to be able to capture all outcomes of interest. These outcomes may not all be present in the data available from individual partners within the DRN. For example, data generated as a result of health care encounters will only show medically-attended events and/or events that generate a bill. In such a situation, a study that utilizes data from within a single inpatient stay would likely find that the DRN provides complete capture of the events of interest. If a heart attack is one of the primary outcomes of a trial, however, the investigator will need to ensure that all events are captured, even the ones where the patient is treated at another facility or dies in the community. Consequently, it may be necessary to supplement DRN data with data about patient status, whether through a call center, a patient portal, data linkage to administrative claims data, or primary data collection via online patient portals, mobile apps, or telephone follow-up.

2.6. Monitoring and reporting

Data for some aspects of data and safety monitoring can come from the CDM, although for site monitoring, the source verification/audit function is typically non-existent within a DRN. There is also a data latency issue when using data from a DRN: partners update their data infrastructure on a regular basis, but data are not in real-time. In some cases, this can be an important aspect for safety monitoring for a data safety and monitoring board or other authorities, depending on the clinical question and intervention being tested. Investigators need to understand how frequently the data are refreshed and the implications for follow-up and monitoring.

2.7. Data analysis

Through the DRN, an investigator can extract a patient-level analysis file from every site so it can be analyzed centrally and meet the centralized data sharing requirements dictated by the trial and sponsor. Alternatively, much of the analysis can be performed in a decentralized fashion using distributed analytical tools (e.g., distributed regression, signal detection, propensity score matching, rates over time) [9–13], or through a hybrid model that pulls back patient-level data that have been summaried (i.e., yes/no indicators for specific medical history or medication concepts). ADAPTABLE is using this hybrid approach, and in this way, investigators will get the "minimum necessary" data to do their analysis, which is in-line with good clinical practice guidelines [14].

2.8. Limitations and challenges

Experience with DRNs is still growing, and healthcare system leadership and clinician champions are also crucial for success. Efforts to minimize burden of interventions in these systems is critical [4]. The issues surrounding consent, disclosure and non-disclosure are complex and are addressed elsewhere [15]. Ensuring data quality and completeness remains an ongoing challenge. DRN-based data curation can foster a culture of continuous improvement, but investigators still need to partner as closely as possible with front-line clinicians to determine what the data actually mean; data are always an incomplete representation of the underlying clinical concepts and are used as a surrogate for clinical phenomena [16]. Before initiating a trial, investigators need to understand how readily available the data are, what data elements are available, how often they are refreshed, and the completeness of data elements, in particular, outcomes of interest. DRNs provide a standardized mechanism for performing such analyses, significantly reducing the burden of engaging individually with potential sites. Even so, a DRN is only as a good as the data held by its partners. Events that are not medically-attended, such as gout flares or suicide attempts, will not be in the EHR or claims data, nor will out-of-hospital death, over-the-counter medication, and community-based

immunizations. Patient-reported outcomes and non-health care determinants of health may not be consistently recorded either. Additional data collection or linkage can ameliorate these issues, as will effective randomization, which helps balance these characteristics in trial populations.

3. Discussion

Although DRNs cannot fulfill every aspect of a clinical trial, they can facilitate certain functions. Perhaps most importantly, they are a vital component for trials that use real-world data to generate real-world evidence. Given their access to larger and more diverse populations, as well as health systems with a variety of care practices, DRN-based trials have the potential to produce more generalized results. Working with a DRN is a collaborative effort between its Coordinating Center, funders, and network partners, which allows questions to be refined based on what is possible. Although any CDM is unlikely to capture everything needed for a trial, ancillary tables or data elements can be developed for use in conjunction with the elements in the model. Ongoing data curation that verifies conformance and quality it is also important, as the CDM is the socket from which all the analytic tools spin. When these tools are developed in a re-useable way, new tools can build on the old and allow for more randomized research to be conducted at lower cost.

Contributor and guarantor information

All authors contributed to the development of the distributed research networks described in this paper and to the ideas presented here. Lesley Curtis and Keith Marsolo drafted the manuscript and will serve as guarantors of the work. The corresponding author attests that all listed authors meet authorship criteria and that no others meeting the criteria have been omitted.

Data sharing statement

We did not create data in the development of this manuscript.

Declaration of competing interest

This work is supported within the National Institutes of Health (NIH) Health Care Systems Research Collaboratory by the NIH Common Fund through cooperative agreement U24AT009676 from the Office of Strategic Coordination within the Office of the NIH Director. The views presented here are solely the responsibility of the authors and do not necessarily represent the official views of the National Institutes of Health.

JB, BH and RP have no conflicts of interest to disclose; BS reports grants from NIH during the conduct of the study; KM reports grants from PCORI, during the conduct of the study; personal fees from Novartis, outside the submitted work; AH reports grants and personal fees from AstraZeneca, grants and personal fees from Bayer, personal fees from Boehringer Ingelheim, grants from American Regent, grants and personal fees from Novartis, grants and personal fees from Merck, grants from Verily, outside the submitted work; LC reports grants from GlaxoSmithKline, grants from Novartis, grants from Boston Scientific, from St. Jude, outside the submitted work; SR reports grants from GSK, outside the submitted work.

There are no other financial relationships with any organisations that might have an interest in the submitted work in the previous three years and no other relationships or activities that could appear to have influenced the submitted work.

References

[1] L.H. Curtis, J. Brown, R. Platt, Four health data networks illustrate the potential for A shared national multipurpose big-data network, Health Aff. 33 (2014) 1178–1186, https://doi.org/10.1377/hlthaff.2014.0121.

[2] U.S. Food and Drug Administration, Framework for FDA's Real-World Evidence Program, 2018. https://www.fda.gov/media/120060/download.

[3] Richard Platt, Pamela Tenaerts, Patrick Archdeacon, Elizabeth Chrischilles, Barbara Evans, Adrian Hernandez, Deven McGraw, Cheryl McMahill-, Cheryl McMahill-Walraven, Marsha Raebel, Kristen Rosati, Developing Approaches to Conducting Randomized Trials Using the Mini-Sentinel Distributed Database, 2014. https://www.sentinelinitiative.org/sites/default/files/Methods/Mini-Sentinel_Methods_CTTI_Developing-Approaches-to-Conducting-Randomized-Trials-Usi.pdf. (Accessed 17 February 2017).

[4] E.B. Larson, C. Tachibana, E. Thompson, G.D. Coronado, L. DeBar, L.M. Dember, S. Honda, S.S. Huang, J.G. Jarvik, C. Nelson, E. Septimus, G. Simon, K.E. Johnson, Trials without tribulations: minimizing the burden of pragmatic research on healthcare systems, Healthcare 4 (3) (2016 September) 138–141, https://doi.org/10.1016/j.hjdsi.2015.07.005.

[5] K.E. Johnson, C. Tachibana, G.D. Coronado, L.M. Dember, R.E. Glasgow, S. S. Huang, P.J. Martin, J. Richards, G. Rosenthal, E. Septimus, G.E. Simon, L. Solberg, J. Suls, E. Thompson, E.B. Larson, A guide to research partnerships for pragmatic clinical trials, BMJ 349 (2014), https://doi.org/10.1136/bmj.g6826 g6826–g6826.

[6] C.B. Granger, S.D. Pokorney, Sentinel IMPACT-AFib: Transforming Pragmatic Clinical Trials Using a Nationwide Distributed Claims Database, 2016. https://www.ctti-clinicaltrials.org/briefing-room/webinars/sentinel-impact-afib-transforming-pragmatic-clinical-trials-using-nationwide.

[7] A.F. Hernandez, R.L. Fleurence, R.L. Rothman, The ADAPTABLE trial and PCORnet: shining light on a new research paradigm, Ann. Intern. Med. 163 (8) (2015) 635–636, https://doi.org/10.7326/M15-1460.

[8] E. Pfaff, A. Lee, R. Bradford, J. Pae, C. Potter, P. Blue, P. Knoepp, K. Thompson, C. L. Roumie, D. Crenshaw, R. Servis, D.A. DeWalt, Recruiting for a pragmatic trial using the electronic health record and patient portal: successes and lessons learned, J. Am. Med. Inform. Assoc. 26 (2019) 44–49, https://doi.org/10.1093/jamia/ocy138.

[9] S. Toh, R. Wellman, R.Y. Coley, C. Horgan, J. Sturtevant, E. Moyneur, C. Janning, R. Pardee, K.J. Coleman, D. Arterburn, K. McTigue, J. Anau, A.J. Cook, Combining distributed regression and propensity scores: a doubly privacy-protecting analytic method for multicenter research, CLEP 10 (2018) 1773–1786, https://doi.org/10.2147/CLEP.S178163.

[10] M. Zhou, S.V. Wang, C.E. Leonard, J.J. Gagne, C. Fuller, C. Hampp, P. Archdeacon, S. Toh, A. Iyer, T.S. Woodworth, E. Cavagnaro, C.A. Panozzo, S. Axtman, R. M. Carnahan, E.A. Chrischilles, S. Hennessy, Sentinel modular program for propensity score–matched cohort analyses: application to glyburide, glipizide, and serious hypoglycemia, Epidemiology 28 (2017) 838–846, https://doi.org/10.1097/EDE.0000000000000709.

[11] Q. Her, Y. Vilk, J. Young, Z. Zilu, J. Malenfant, S. Malek, D. Toh, ICPE Presentation: Performance of a Distributed Regression Analysis Software and Workflow, 2018, p. 2018. https://www.sentinelinitiative.org/communications/publications/2018-icpe-presentation-performance-distributed-regression-analysis.

[12] S. Toh, M.E. Reichman, M. Houstoun, X. Ding, B.H. Fireman, E. Gravel, M. Levenson, L. Li, E. Moyneur, A. Shoaibi, G. Zornberg, S. Hennessy, Multivariable confounding adjustment in distributed data networks without sharing of patient-level data: confounding Adjustment in Distributed Data Networks, Pharmacoepidemiol. Drug Saf. 22 (2013) 1171–1177, https://doi.org/10.1002/pds.3483.

[13] J. Maro, M. Nguyen, R. Izem, Y.-H. Hsueh, I. Dashevsky, A. Cosgrove, C. Jones, J. Major, E. Zhou, E. Baro, J. Gagne, S. Wang, M. Kulldorff, ICPE Presentation: Signal Detection Using TreeScan with Drug Classes: Pilot Projects in Sentinel, 2018, p. 2018. https://www.sentinelinitiative.org/communications/publications/2018-icpe-presentation-signal-detection-using-treescan-drug-classes. (Accessed 11 December 2019).

[14] U.S. Food and Drug Administration, FDA Regulations Relating to Good Clinical Practice and Clinical Trials, 2015 (accessed June 16, 2015), http://www.fda.gov/ScienceResearch/SpecialTopics/RunningClinicalTrials/ucm155713.htm.

[15] K.P. Weinfurt, J. Sugarman, Consent, disclosure, and non-disclosure, in: Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials, 2017. http://www.rethinkingclinicaltrials.org/consent-disclosure-non-disclosure-top/consent-disclosure-and-non-disclosure-introduction/.

[16] R.L. Richesson, R. Platt, G.E. Simon, L.H. Curtis, R.L. Laws, A.F. Hernandez, J. Puro, D.F. Zatzick, Using electronic health record data in pragmatic clinical trials: data as a surrogate for clinical phenomena, in: Rethinking Clinical Trials: A Living Textbook of Pragmatic Clinical Trials., NIH Health Care Systems Research Collaboratory, 2017. http://www.rethinkingclinicaltrials.org/using-electronic-health-record-data-pragmatic-clinical-trials-top/6735-2/. (Accessed 8 September 2017).