Brief Communication

Accrual to Clinical Trials (ACT): A Clinical and Translational Science Award Consortium Network

Shyam Visweswaran,1 Michael J. Becich,1 Vincent S. D’Itri,2 Elaina R. Sendro,2 Douglas MacFadden,3 Nicholas R. Anderson,4 Karen A. Allen,5 Dipti Ranganathan,6 Shawn N. Murphy,7 Elaine H. Morrato,8 Harold A. Pincus,9 Robert Toto,10 Gary S. Firestein,11 Lee M. Nadler,3 and Steven E. Reis12

1Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, Pennsylvania, USA, 2The Chartis Group, Chicago, Illinois, USA, 3The Harvard Clinical and Translational Science Center, Harvard Medical School, Boston, Massachusetts, USA, 4Department of Public Health Sciences, University of California, Davis, California, USA, 5Office of Research, University of California, Irvine, California, USA, 6Academic Information Systems, University of Texas Southwestern Medical Center, Dallas, Texas, USA, 7Research Computing, Partners HealthCare, Charlestown, Massachusetts, USA, 8Department of Health Systems, Management and Policy, Colorado School of Public Health, University of Colorado Anschutz Medical Campus, Aurora, Colorado, USA, 9Department of Psychiatry, Columbia University, New York, New York, USA, 10The Center for Translational Medicine, University of Texas Southwestern Medical Center, Dallas, Texas, USA, 11Altman Clinical and Translational Research Institute, University of California, San Diego, California, USA, and 12The Clinical and Translational Science Institute, University of Pittsburgh, Pittsburgh, Pennsylvania, USA

Corresponding Author: Shyam Visweswaran, MD, PhD, Department of Biomedical Informatics, University of Pittsburgh, Pittsburgh, The Offices at Baum, 5607 Baum Blvd., Suite 523, Pittsburgh, PA 15206, USA (shv3@pitt.edu).

Received 27 April 2018; Revised 15 June 2018; Accepted 13 July 2018

ABSTRACT

The Accrual to Clinical Trials (ACT) network is a federated network of sites from the National Clinical and Translational Science Award (CTSA) Consortium that has been created to significantly increase participant accrual to multi-site clinical trials. The ACT network represents an unprecedented collaboration among diverse CTSA sites. The network has created governance and regulatory frameworks and a common data model to harmonize electronic health record (EHR) data, and deployed a set of Informatics for Integrating Biology and the Bedside (i2b2) data repositories that are linked by the Shared Health Research Information Network (SHRINE) platform. It provides investigators the ability to query the network in real time and to obtain aggregate counts of patients who meet clinical trial inclusion and exclusion criteria from sites across the United States. The ACT network infrastructure provides a basis for cohort discovery and for developing new informatics tools to identify and recruit participants for multi-site clinical trials.

Key words: clinical trials, accrual, cohort discovery, clinical data research network, electronic health records

INTRODUCTION

Advancing translational research to improve human health requires that sufficient numbers of participants are available for clinical investigations. It is therefore critical to efficiently identify eligible participants, provide them with research opportunities, and enroll them in research studies. These tasks are particularly challenging for multi-site clinical trials, the majority of which are unable to recruit their proposed number of participants within their planned time frame.1–4 Failure to accrue eligible participants in trials is inefficient and wasteful, limits generalizability of results, and leads to premature closure of trials.5
Electronic health records (EHRs) are now used clinically at nearly all major medical centers and hospitals. EHR data provide an opportunity to significantly increase the efficiency of clinical trials by identifying eligible participants based on demographic, diagnostic, procedural, laboratory, medication, and other information. However, there are important privacy and regulatory concerns that need to be addressed in order to use the EHR data cooperatively. In addition, differences in data representation across institutional EHRs pose technical challenges in harmonizing and sharing data.

THE ACCRUAL TO CLINICAL TRIALS NETWORK

To address these challenges, four Clinical and Translational Science Award (CTSA) sites collaborated to create the Accrual to Clinical Trials (ACT) network supported by funding from the National Center for Advancing Translational Sciences (NCATS). The network is expanding its reach by adding new CTSA sites in waves and is enhancing the functionality of the network in three stages. Stage I will enable rapid real-time cohort exploration across the federated network. Stage II will enable identification and contact of participants who are eligible for clinical trial recruitment. Stage III will develop approaches, tools, educational materials, and infrastructure to enable patients and care providers to identify clinical trials.

The ACT network has developed a governance structure, constructed a regulatory framework, deployed technical infrastructure, created a common data model (CDM), and linked CTSA sites to enable rapid cohort exploration. Institutional leaders at participating sites signed a data sharing agreement and the ACT protocol was approved by local Institutional Review Boards (IRBs). The network links 21 CTSA sites and has implemented Stage I functionality to enable cohort discovery across more than 40 million patients for clinical trials as well as for other cohort studies.

The ACT network complements the National Patient-Centered Clinical Research Network (PCORnet) that is funded by the Patient-Centered Outcomes Research Institute (PCORI). The goal of PCORnet is to create a network of clinical data research networks (CDRNs) that leverage EHR data for comparative effectiveness and other types of research. The goal of the ACT network is to enable efficient, safe and lower cost multi-site clinical trials and translational research studies for all CTSAs.

ORGANIZATION

The ACT network organization consists of a group of four Principal Investigators (PIs), an executive committee, five work groups, staff at each site including a project manager, technical and dissemination personnel, and a central project management team (see Figure 1). The PI group is responsible for the strategic goals, benchmarks, metrics, and timely implementation. The executive committee is formed by the PI group, work group leads, the central project management team, and a NCATS program officer. The committee functions to optimize all operational decisions and decides on the priorities and deliverables for each work group.

The work group leads are responsible for scheduling work group meetings, implementation of their assigned tasks, recording the completion of milestones and facilitating face-to-face meetings to ensure integration, collaboration, and meeting benchmarks. The governance work group is responsible for defining the structure of ACT operations, including the creation of a governance process. ACT governance attempts to strike a balance between ensuring an
PARTICIPATING SITES

At present, the ACT network brings together 21 CTSA sites from 16 U.S. states and the District of Columbia and includes several pediatric academic health centers (see Figure 2). In the future, additional CTSA sites will be invited to join the network, and by project completion, up to 64 sites will be connected. The current ACT network contains data on more than 40 million patients (see Tables 1 and 2). Each site contributed at a minimum data from January 2012 though many sites have contributed far more.

INFORMATICS INFRASTRUCTURE

Technology

The network consists of local Informatics for Integrating Biology at the Bedside (i2b2) EHR data repositories that are integrated by the Network consists of local Informatics for Integrating Biology and Technology (i2b2) EHR data repositories. The network consists of local Informatics for Integrating Biology at the Bedside (i2b2) EHR data repositories that are integrated by the Network consists of local Informatics for Integrating Biology at the Bedside (i2b2) EHR data repositories.

For operational efficiency, there are three separate networks that include test, stage, and production networks. The test network, comprising of four sites, is used for testing software and ontology upgrades. The stage network consists of new sites that are connecting to the network for the first time and provides a mechanism to evaluate a site’s readiness and troubleshoot technical, ontology, and data issues. The production network is used by investigators for cohort exploration.

Data model and i2b2/SHRINE ontologies

The ACT CDM specifies data domains and data elements to be loaded at each site’s i2b2 repository. The domains in the current ACT CDM include demographics, diagnoses, procedures, medications, laboratory test results, and visit characteristics; these comprise a subset of the domains in the PCORnet CDM. For each data domain such as diagnoses or procedures, the ACT CDM (version 1.4) has fewer data elements when compared with the corresponding data domain in the PCORnet CDM (version 4.1); however, a data element that is included in the ACT CDM has the same definition as
Table 2. The Accrual to Clinical Trials network demographics

<table>
<thead>
<tr>
<th>Age (years)</th>
<th>Number of patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>0–9</td>
<td>3,262,900 (8)</td>
</tr>
<tr>
<td>10–17</td>
<td>3,065,800 (8)</td>
</tr>
<tr>
<td>18–34</td>
<td>7,377,500 (19)</td>
</tr>
<tr>
<td>35–44</td>
<td>4,866,700 (12)</td>
</tr>
<tr>
<td>45–54</td>
<td>5,371,600 (14)</td>
</tr>
<tr>
<td>55–64</td>
<td>5,477,200 (14)</td>
</tr>
<tr>
<td>65–74</td>
<td>4,233,900 (11)</td>
</tr>
<tr>
<td>75–84</td>
<td>2,456,200 (6)</td>
</tr>
<tr>
<td>85–90</td>
<td>708,700 (2)</td>
</tr>
<tr>
<td>≥90</td>
<td>2,224,300 (6)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Race</th>
<th>Number of patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asian</td>
<td>941,000 (4)</td>
</tr>
<tr>
<td>Black or African American</td>
<td>3,276,700 (16)</td>
</tr>
<tr>
<td>American Indian or Alaskan Native</td>
<td>119,200 (0.6)</td>
</tr>
<tr>
<td>Native Hawaiian or Other Pacific Islander</td>
<td>175,500 (0.8)</td>
</tr>
<tr>
<td>White</td>
<td>16,533,000 (79)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Ethnicity</th>
<th>Number of patients (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hispanic</td>
<td>2,421,000 (14)</td>
</tr>
<tr>
<td>Not hispanic</td>
<td>37,773,100 (86)</td>
</tr>
</tbody>
</table>

The corresponding data element in PCORnet CDM. The ACT CDM is available from the ACT website at http://www.actnetwork.us.

SHRINE, as does i2b2, employs a query language that allows querying of data across the sites in the network using predefined terms. For each data domain (such as diagnoses or medications) the predefined collection of terms that describe the data for the domain are arranged in a hierarchy for easy navigation. These ontologies are developed centrally and distributed for installation at each ACT site and are available from the ACT wiki (https://ncatswiki.dbmi.pitt.edu). The ACT ontologies are updated periodically to keep up with expansion of data domains, changes and updates in source terminologies, and to fix errors found in the deployed ontologies.

Data updates and data characterization

To ensure recency, each site updates its i2b2 repository monthly. To ensure data quality, development and deployment of data characterization processes is ongoing. Guidance for extract-transform-load process including recommendations for the types of data and sources of data and i2b2 and ontology specific formats is provided through regular conference calls. An online data characterization survey has been implemented that will be completed by each site. Data from this survey will provide site-specific information on the time period of data, amount of data along several dimensions, expected gaps in data (eg, a children’s hospital may not have patients older than 18 years of age), and results of checks on data formats in the i2b2 repository.

Test cases

Several high-priority clinical trials were employed as test cases for preliminary evaluation of the network. Examples of clinical trials that were used as test cases included identification of patients with early rheumatoid arthritis and an inadequate response to methotrexate, identification of patients with early stage fibrosis secondary to Hepatitis C infection, and identification of patients with coronary artery disease who are eligible for chelation therapy. For each test case, inclusion and exclusion criteria of the corresponding clinical trial was translated into a SHRINE query using the ACT ontologies, and the query was executed from the University of Pittsburgh site. Since ACT uses a hub and spoke network topology, the query that is issued from the University of Pittsburgh site is transmitted to the SHRINE hub that is located at the Harvard Medical School from where it is broadcast to all network sites. Counts resulting from the execution of the query at each site’s i2b2 repository are transmitted to the site that issued the query via the SHRINE hub. Queries often execute in 5 min or under and the investigator is provided with a list of sites on the network and corresponding counts. As an example, the inclusion and exclusion criteria for the chelation therapy test case and coverage for these criteria in the ACT ontologies are shown in Supplementary Appendix Tables S1 and S2 respectively. This query returned counts from 21 sites on the network in 5 min and identified a total of 8383 patients across the network.

Challenges

The ACT network is a large, real-time and self-serve network and these features pose several challenges. Maintaining continuous connectivity of every site to the network is exacting. Currently, network status is assessed weekly using a “smoke test” that interrogates every data domain based on which a report is circulated that provides technical and connectivity status of each site’s i2b2 repository (see Figure 3). A helpdesk and a mailing list provide guidance and assistance with troubleshooting. Update of software or ontologies have to be done in a narrow maintenance window so that the network is available for querying for the maximum length of time. In addition, in a large network, it is useful to be able to issue a series of queries without waiting for each query to complete at each site. A future version of SHRINE will include this capability. The source terminologies that are used to develop the ontologies typically retire codes. However, to maintain ability to query historical data the ontologies have to include retired codes, and the process of obtaining retired codes can be challenging depending on the terminology. Ensuring the quality of the data at each site is critical for large scale use of the network; this is particularly challenging since data characterization needs to be performed with every update of the data repository at each site. Work is ongoing in developing data characterization processes that can be efficiently deployed across the network.

SUMMARY

The long-term goal of the ACT network is to transform clinical and translational research by developing an efficient and extensible digital infrastructure that enables CTSA sites to collaborate in cohort exploration, to identify and contact patients who are eligible for clinical trials, and to enable patients and care providers to identify clinical trials. Currently, the ACT network enables cohort exploration by providing aggregate counts from more than 40 million patients who meet clinical trial inclusion criteria across 21 sites. In the future, the ACT network will attempt to extend to all CTSA sites, add additional data domains and data elements to the ACT CDM, and use the ACT infrastructure to develop and deploy novel recruitment tools to enable investigators to efficiently recruit participants across institutions. The ACT network represents an unprecedented collaboration among diverse CTSA sites and serves as a national resource for accelerating recruitment of research participants.
FUNDING

This work was supported by the National Center for Advancing Translational Sciences of the National Institutes of Health under grant numbers UL1 TR000005-09S1, UL1 TR001857-01S1, UL1 TR001876, UL1 TR001873, UL1 TR002553, UL1 TR002378, UL1 TR002541, UL1 TR002529, UL1 TR001450, UL1 TR001422, UL1 TR002369, UL1 TR001085, UL1 TR001860, UL1 TR001414, UL1 TR001881, UL1 TR001442, UL1 TR001872, UL1 TR001425, UL1 TR002535, UL1 TR001427, UL1 TR002494, UL1 TR001857, and UL1 TR001105.

CONTRIBUTORS

S.V. conceived and designed the study, participated in data collection, analysis, and interpretation, drafted and revised the manuscript and approved the final version for submission. M.J.B. made critical manuscript revisions and approved the final version for submission. V.S.D. participated in data collection, made critical manuscript revisions, and approved the final version for submission. E.R.S participated in data collection and approved the final version for submission. D.M. participated in data collection, drafted and revised the manuscript and approved the final version for submission. N.R.A. drafted and revised the manuscript and approved the final version for submission. K.A.A. drafted and revised the manuscript and approved the final version for submission. D.R. made critical manuscript revisions and approved the final version for submission. G.S.F. obtained funding, made critical manuscript revisions and approved the final version for submission. L.M.N. obtained funding, made critical manuscript revisions and approved the final version for submission. S.E.R. obtained funding, made critical manuscript revisions and approved the final version for submission.

SUPPLEMENTARY MATERIAL

Supplementary material is available at Journal of the American Medical Informatics Association online.

Conflict of interest statement. None declared.

REFERENCES


