AMIA
INFORMATICS PROFESSIONALS. LEADING THE WAY.

OXFORD

Research and Applications

EHR-based cohort assessment for multicenter RCTs: a fast and flexible model for identifying potential study sites

Sarah J. Nelson[1], Bethany Drury[1], Daniel Hood[2], Jeremy Harper[2], Tiffany Bernard[3], Chunhua Weng[4], Nan Kennedy[1], Bernie LaSalle[5], Ramkiran Gouripeddi [2], Consuelo H. Wilkins[1,6,7,8], and Paul Harris[1,9]

[1]Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [2]Regenstrief Institute, Indianapolis, Indiana, USA, [3]The Ohio State University Wexner Medical Center, Columbus, Ohio, USA, [4]Department of Biomedical Informatics, Columbia University, New York, New York, USA, [5]University of Utah, Salt Lake City, Utah, USA, [6]Department of Medicine, Vanderbilt University Medical Center, Nashville, Tennessee, USA, [7]Department of Internal Medicine, Meharry Medical College, Nashville, Tennessee, USA, [8]Office of Health Equity, Vanderbilt University Medical Center, Nashville, Tennessee, USA, and [9]Department of Biomedical Informatics, Vanderbilt University Medical Center, Nashville, Tennessee, USA

Corresponding Author: Paul A. Harris, PhD, Vanderbilt Institute for Clinical and Translational Research, Vanderbilt University Medical Center, 2525 West End Avenue, Suite 1500, Nashville, TN 37203, USA; paul.harris@vumc.org

ABSTRACT

Objective: The Recruitment Innovation Center (RIC), partnering with the Trial Innovation Network and institutions in the National Institutes of Health-sponsored Clinical and Translational Science Awards (CTSA) Program, aimed to develop a service line to retrieve study population estimates from electronic health record (EHR) systems for use in selecting enrollment sites for multicenter clinical trials. Our goal was to create and field-test a low burden, low tech, and high-yield method.

Materials and Methods: In building this service line, the RIC strove to complement, rather than replace, CTSA hubs' existing cohort assessment tools. For each new EHR cohort request, we work with the investigator to develop a computable phenotype algorithm that targets the desired population. CTSA hubs run the phenotype query and return results using a standardized survey. We provide a comprehensive report to the investigator to assist in study site selection.

Results: From 2017 to 2020, the RIC developed and socialized 36 phenotype-dependent cohort requests on behalf of investigators. The average response rate to these requests was 73%.

Discussion: Achieving enrollment goals in a multicenter clinical trial requires that researchers identify study sites that will provide sufficient enrollment. The fast and flexible method the RIC has developed, with CTSA feedback, allows hubs to query their EHR using a generalizable, vetted phenotype algorithm to produce reliable counts of potentially eligible study participants.

Conclusion: The RIC's EHR cohort assessment process for evaluating sites for multicenter trials has been shown to be efficient and helpful. The model may be replicated for use by other programs.

Key words: medical informatics, electronic health records, multicenter studies as topic, cohort assessment, randomized controlled trials as topic

INTRODUCTION

Clinical and translational research contributes significantly to the health and welfare of society. Although research across the spectrum of clinical trial methodologies is valuable, the Coronavirus Disease of 2019 (COVID-19) pandemic has served to highlight the importance of randomized controlled trials (RCTs). Only through the systematic evaluation of new treatments via well-designed RCTs can evidence of safety and efficacy be established.[1] Under-enrollment in RCTs, however, is a persistent problem.[2] Large-scale RCTs often require multiple enrollment sites to obtain a volume of participants necessary to detect significant outcome signals and to ensure that sample diversity is adequate to fully represent all segments of society in study findings.[3,4]

In 2006, The National Institutes of Health created the Clinical and Translational Science Awards (CTSA) program,[5] in part, to address this very issue. By fostering collaboration on a national scale among CTSA hub institutions, research and discovery of new treatments have been accelerated.[5] The Trial Innovation Network (TIN),[6] which includes the CTSA hubs, was instituted to focus specifically on operational innovation, efficiency, and excellence in conducting clinical trials. In partnership with the TIN, the Recruitment Innovation Center (RIC) was established as an evidence-based center that works directly with research teams to develop, demonstrate, and disseminate novel methods, strategies, and tools to increase participant recruitment and retention.

A critical element of the RIC's mission is to facilitate collaboration among CTSA hubs to amplify recruitment potential for multicenter RCTs. One of the RIC's early goals included building a service line designed to assist RCT principal investigators (PIs) in retrieving study population estimates from CTSA Network hubs and affiliates for use in evaluating sites for potential collaboration in multicenter studies. Our goal has been to create, and field-test a fast and flexible system that is low burden, low tech, and high yield. We present details of our service line and approach here along with how it has evolved since exercising our model across 36 expressions of interest requests that included an electronic health record (EHR)-based cohort assessment component between 2017 and 2020. Our experience and lessons learned are not limited to a single informatics platform, data model, or biomedical science disease domain. Given this, we believe our model and experiential findings will provide value to the larger clinical trial informatics communities.

MATERIALS AND METHODS

Creating a model

We started with a set of basic assumptions: (1) all CTSA hubs and affiliate sites have some means of leveraging a local research data warehouse to perform cohort identification queries; (2) individual hub capabilities likely include a diverse set of options for self-service cohort count estimates (eg, i2b2,[7] Leaf,[8] ATLAS,[9] TriNetX,[10] and Accrual to Clinical Trials [ACT][11]) as well as data engineers who have access to back-end data sources if needed for difficult query logic not possible using self-service user interface tools; (3) underlying data models and data availability and latency (eg, types, temporality) will vary across hubs; and (4) the development and implementation of effective EHR-based phenotyping algorithms are often more complex than most RCT PIs realize unless they have prior experience.

We also started with a core set of guiding principles: (1) respect the autonomy, diversity, and capacity of any single hub by building a process that promotes inclusivity and ability to participate in all desired cohort query exercises; (2) create minimum burden for hubs desiring to compute and report local patient count queries; and (3) complement, rather than compete with, existing self-service cohort generator platforms such as ACT, PCORnet Clinical Data Research Networks,[12] and TriNetX.

Creating a process

TIN studies requesting a multicenter EHR-based cohort assessment are assigned a RIC consultant who works with the study PI to develop a computable phenotyping algorithm representative of the desired study population. These consultants are trained research professionals who are supported by informaticians and data scientists familiar with EHR-derived phenotypes. Each phenotype is unique and developed with 3 main considerations in mind: (1) data that are readily available in the EHR; (2) a phenotype that is generalizable enough for hubs to run on diverse local data warehousing architecture and self-service tools; and (3) a phenotype that can be run in a reasonable amount of time with minimal burden at each site.

Based on these guiding considerations, an RIC consultant works with the investigator to translate criteria necessary to identify patient eligibility into standardized concepts which can be derived from a local EHR system. The goal is to create a phenotype algorithm that answers the underlying question, "How is this study's disease/condition represented across the CTSA Consortium?" When possible, the algorithm contains equivalent codes (eg, International Classification of Diseases—ICD-9/ICD-10, Logical Observation Identifiers Names and Codes/Current Procedural Terminology) so that hubs have the flexibility to use the code system which aligns with their underlying source data. Phenotype algorithm creation is an iterative process, usually consisting of 1 to 2 meetings with the investigator. Once the PI approves the final algorithm version, usually within a week, the RIC informatics resource lead will initiate the vetting process to answer any remaining questions before sending out to the larger CTSA network to run at each institution.

The RIC phenotype vetting process begins at Vanderbilt University Medical Center (VUMC), where the algorithm is reviewed by a senior team of program managers and analysts. If there are any questions or errors noticed from the VUMC reviewers, the resource lead will contact the investigator for clarification or suggested changes. Based on the changes, the phenotype may need to be reviewed again at VUMC before the request for review from RIC partner institutions (Columbia, Regenstrief, Utah, and Ohio State University [OSU]) is initiated. If no or only minor revisions that do not change the anticipated results of the algorithm are needed, the RIC resource lead will run the query at VUMC and distribute the final approved version of the algorithm to the RIC partner institutions. VUMC uses a custom-built in-house system called Record Counter.[13] The RIC partner institutions complete the request by running on their own platforms and returning query counts, any feedback, and a calculation of the time required to run the query. These institutions vary in how they obtain their counts. Columbia University runs these queries manually using data programmers with direct access to research data warehouse infrastructure. Regenstrief Institute and The OSU run their phenotype queries using i2b2. The University of Utah uses TriNetX. Collectively, these platforms represent common query tools being used across CTSA hubs, which along with heterogeneity of data warehouse practices across institutions further establishes the generalizability of the phenotype.

The RIC consultant shares the phenotype vetting results with the investigator to confirm that the returned counts are within the expected ranges. Figure 1 shows the RIC-developed phenotype development and vetting process.

During the phenotype vetting process, we ask the 4 RIC partner institutions to indicate how long the phenotype took to run on each of their respective query systems. We add 1 h to this time to account for presumed involvement by a local project analyst, and then include this total time as an estimated burden in messaging to CTSA hubs during the broader expression of interest request. Based on feedback from hubs, the RIC has established a goal of <4 h for hub completion when developing and disseminating phenotype count requests. To evaluate the accuracy of our approach, the RIC includes a question on the cohort assessment response survey that asks each hub to indicate burden for the query process.

To promote inclusivity for CTSA hubs, phenotype queries were developed using a tiered approach. First-tier phenotyping started with basic EHR data types and subsequent tiers included additional date types and complex relationships. This approach typically allowed all interested hubs to return results for at least some portion of the phenotype request. The hub results survey included a dedicated space to add comments supporting their submission.

The RIC distributes the phenotyping algorithm to hubs in a human-mediated process with simple instructions for running the algorithm locally with a RIC-generated survey in REDCap (Research Electronic Data Capture)[14] that is designed to collect and collate hub results. The RIC then generates a report for communication with the study PI to be used in the trial site selection process. Prior to sharing the report with the investigator, EHR cohort assessment information is reviewed in aggregate to identify and flag obvious outliers (usually attributed to a simple algorithm application error—eg, reversing an "include" statement with an "exclude statement"). This aggregated quality check is generally low burden for the RIC and usually remedied by communication and subsequent re-run by the outlier site. The comprehensive report includes aggregate count responses from each hub wishing to participate as a trial site, visual charts representing the cohort assessment submissions from each prospective site individually as well as a visualization of all count responses combined. Only responding hubs are included in this report. This includes hubs that submitted cohort counts or opted out. If provided by the site, the report also includes the reason for opt out.

In 2019, the RIC began to leverage the ACT and TriNetX platforms to run algorithms prior to sending the request to the CTSA hubs to run at their own institutions. We are now able to run queries, when compatible, through these networks and share site-specific prerun counts. Hubs are then able to review the results from ACT and/or TriNetX and choose to submit the counts as their institution's response, or they may choose to run the query through an analyst at their site instead. Figure 2 shows the full EHR-based cohort assessment process.
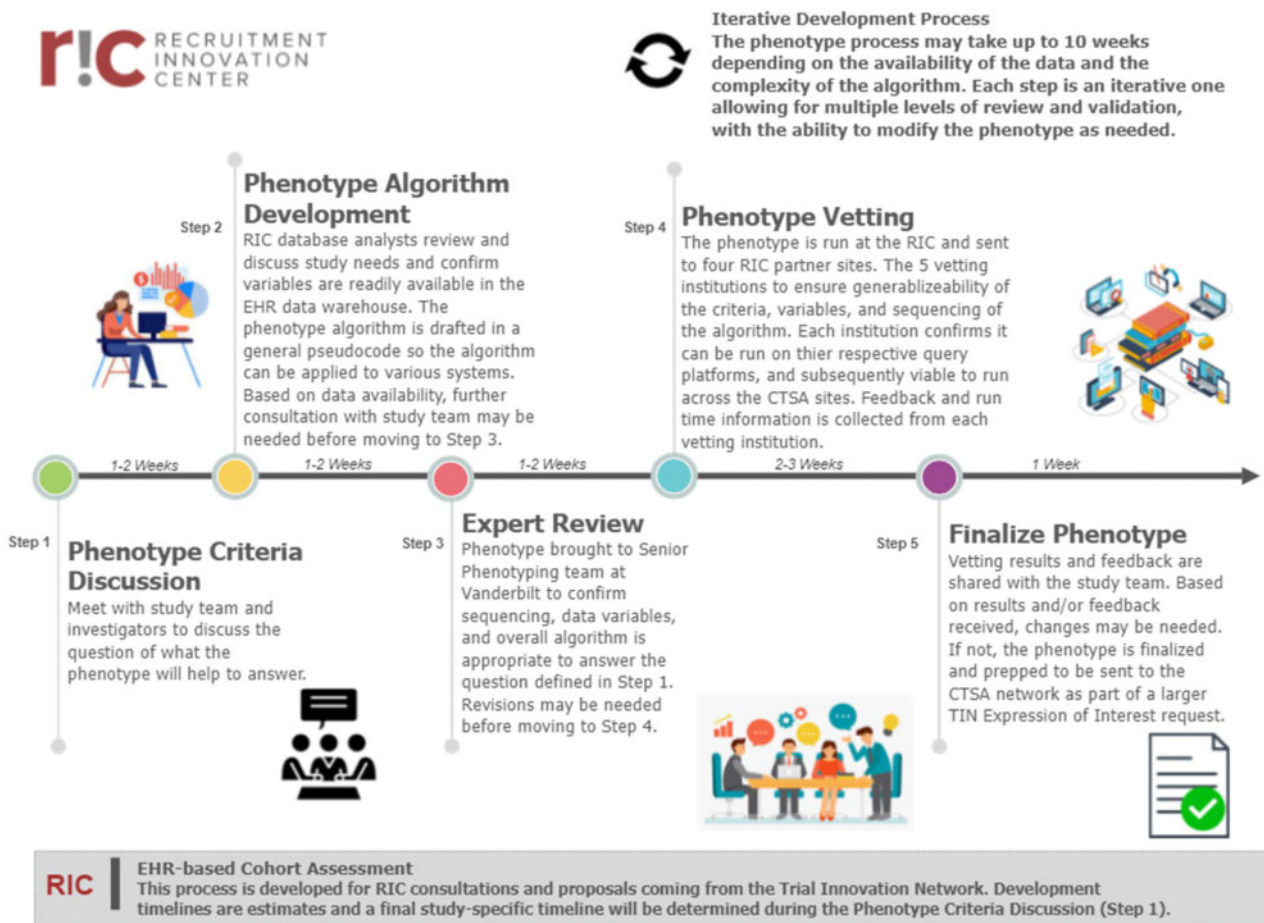

Iterative Development Process
The phenotype process may take up to 10 weeks depending on the availability of the data and the complexity of the algorithm. Each step is an iterative one allowing for multiple levels of review and validation, with the ability to modify the phenotype as needed.

Step 1 — Phenotype Criteria Discussion
Meet with study team and investigators to discuss the question of what the phenotype will help to answer.

1-2 Weeks

Step 2 — Phenotype Algorithm Development
RIC database analysts review and discuss study needs and confirm variables are readily available in the EHR data warehouse. The phenotype algorithm is drafted in a general pseudocode so the algorithm can be applied to various systems. Based on data availability, further consultation with study team may be needed before moving to Step 3.

1-2 Weeks

Step 3 — Expert Review
Phenotype brought to Senior Phenotyping team at Vanderbilt to confirm sequencing, data variables, and overall algorithm is appropriate to answer the question defined in Step 1. Revisions may be needed before moving to Step 4.

1-2 Weeks

Step 4 — Phenotype Vetting
The phenotype is run at the RIC and sent to four RIC partner sites. The 5 vetting institutions to ensure generalizeability of the criteria, variables, and sequencing of the algorithm. Each institution confirms it can be run on thier respective query platforms, and subsequently viable to run across the CTSA sites. Feedback and run time information is collected from each vetting institution.

2-3 Weeks

Step 5 — Finalize Phenotype
Vetting results and feedback are shared with the study team. Based on results and/or feedback received, changes may be needed. If not, the phenotype is finalized and prepped to be sent to the CTSA network as part of a larger TIN Expression of Interest request.

1 Week

RIC | EHR-based Cohort Assessment
This process is developed for RIC consultations and proposals coming from the Trial Innovation Network. Development timelines are estimates and a final study-specific timeline will be determined during the Phenotype Criteria Discussion (Step 1).

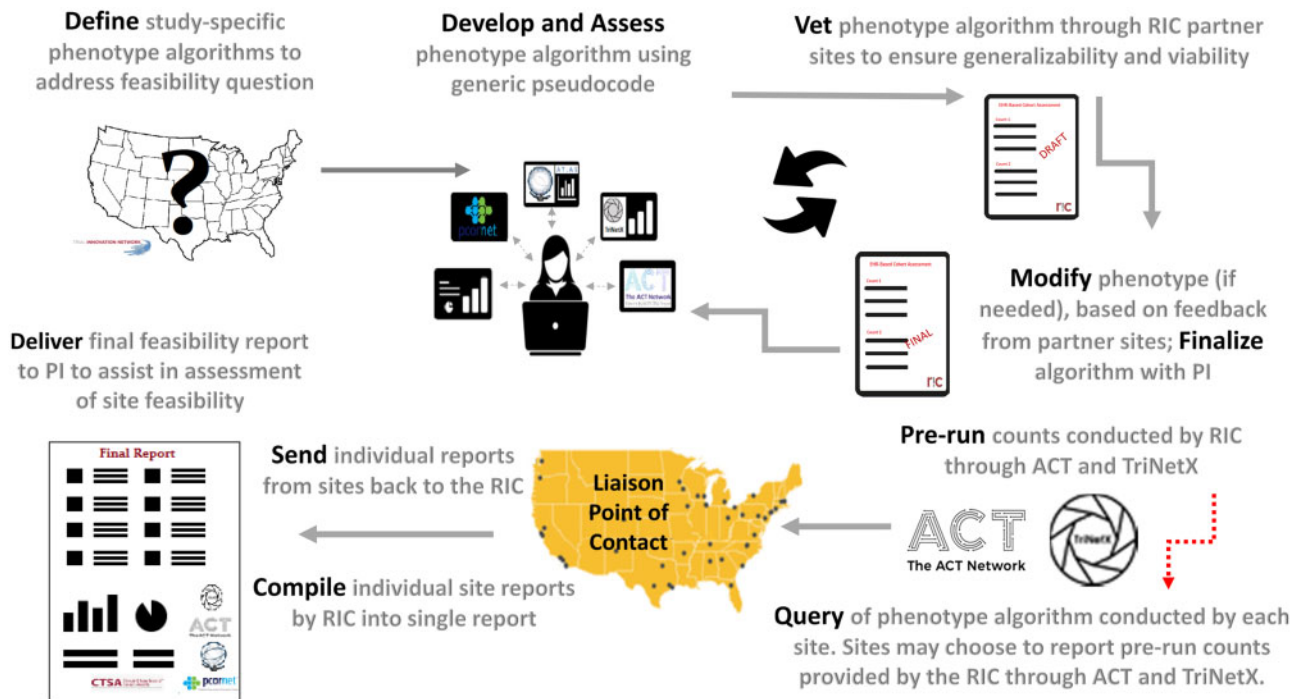**Figure 1.** Recruitment Innovation Center phenotype development and vetting process.

Figure 2. "Recipe" for the electronic health record Cohort Assessment model.

Since the creation of this resource line, we have continually evolved our process to streamline and minimize the burden for CTSA hubs when responding to EHR cohort assessment requests. We encourage feedback on our process and can implement and share changes through regular communication. We have created a designated location on the TIN intranet site (TIN Dashboard) which supports communication about all past and current projects considered and supported by the network. Liaisons from all CTSA hubs can view information about the cohort assessment request, such as funding status, request due date, hub-specific results, and a study timeline. Figure 3 shows the EHR Cohort Assessment section of the TIN Dashboard in context of a single trial. Hub liaisons and designees can communicate with cohort assessment project leaders directly through the Dashboard using the "contact us" feature. The "Discussion Forum" is an interactive space where questions can be posted for any other user to answer. The most common questions posted are regarding mapping of specific code types. This user forum allows informatics experts from any CTSA hub to share their advice and tips on successfully running a query. Users can subscribe to the entire forum or specific postings.

RESULTS

Between August 2017 and December 2020, the RIC developed and socialized 36 phenotype-dependent cohort count requests in support of RCTs desiring new enrollment sites. Table 1 shows detailed response metrics, query information, and phenotype variables for cohort assessment requests between mid-2018 and 2020. The 4 factors with the greatest impact on runtime are (1) the total number of counts being requested in the query; (2) requests to retrieve problematic information such as insurance type or payer information; (3) layering 2 or more variable types (diagnosis code, procedure code, lab, etc.) in 1 count.; and (4) presence of and complexity of temporal requirements (eg, observation A occurs within X days of observation B and Y days of observation C).

All CTSA hubs received and responded to at least 1 RCT cohort assessment request. Across all CTSA hubs the average response rate was 73% with an average response time of 28 days. Responses from the hubs can be either a submission of cohort counts or an opt out. On average, 23% of responses received were an opt out, meaning the CTSA hub declined to participate in the trial as a site. Common reasons for opting out are the inability to identify a local investigator, local informatics capacity, or no interest in the study. In mid-2018, hubs began reporting the amount of time it took to run each cohort assessment. Runtime data were received for 30 queries, 9 of these averaged 4 or more hours. The response rate for these 9 "long runtime queries" was 73%. Thus, while queries were taking longer to run, hubs consistently responded to the requests when they were interested in participation as a potential site.

To minimize burden on CTSA hubs, the RIC-developed processes to provide pre-run counts from either ACT or TriNetX in a total of 7 studies (7 TriNetX and 3 ACT). Across the 7 studies, a total of 47 hubs submitted prerun TriNetX counts while 9 submitted prerun ACT counts. To collect this information, a question was added to the response survey to indicate whether the counts being submitting were RIC-provided prerun counts from either of these platforms. We did not find any 1 hub consistently submitting partial results, nor did we find hubs using ACT, i2b2, or TriNetX any more likely to return partial results than other hubs.

In addition to assessing burden at individual hubs, we tracked RIC effort to generate, disseminate, and aggregate phenotype across all hubs. Effort includes consultative work and meetings, phenotype development and review, facilitating and conducting vetting of the phenotype algorithms, as well as communication with hubs during each outreach request. On average, each supported study requires about 15 h of employee effort. Like runtime reported by sites, complexity of the phenotype was increased the burden and work re-
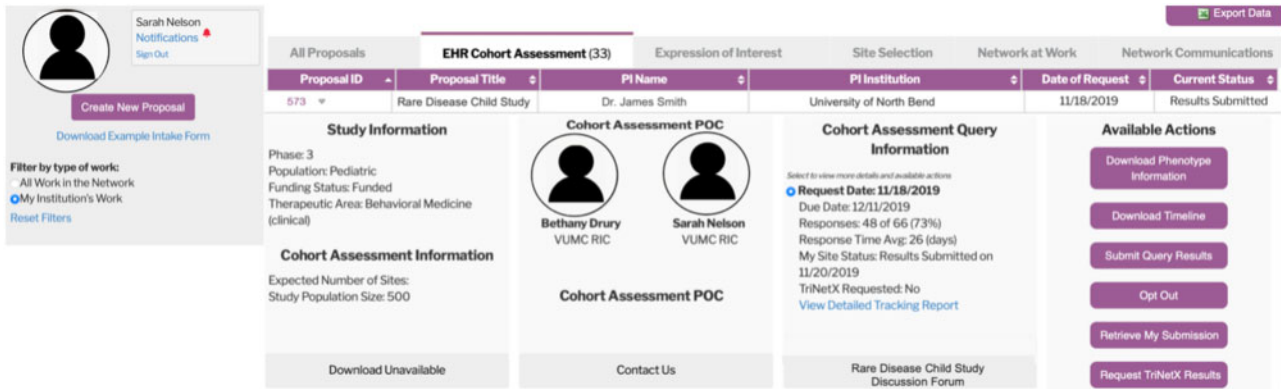
Figure 3. EHR cohort assessment section of Trial Innovation Network dashboard.

quired for the RIC consultant. Other factors that impact RIC effort are length of algorithm and engagement behavior of lead investigator teams.

DISCUSSION

Planning a successful multicenter clinical trial requires that researchers identify study sites that can find, recruit, and enroll the projected number of participants needed to complete the trial. A first step is for researchers to calculate the total number of participants that will be adequate to show statistical significance in the study intervention. Next, investigators must determine the approximate number of participants likely to enroll at individual sites to ascertain the number of sites required. Refusal and dropout rates must be factored into the equation, as well as extra recruitment efforts to enroll a sufficiently diverse participant population.

Investigators essentially perform a balancing act between their estimates of participant enrollment and the number of study sites engaged. Although more sites mean a larger potential participant pool, each additional site increases time and cost. Study start-up requirements at each site include contract execution, local IRB approval (if a central IRB is not in place), assessment of the impact of competing trials, and additional staff training. Investigators must adequately budget for each site and are limited by their funding.

Finalizing study sites can be a difficult decision for investigators and many factors must be considered upfront to avoid mid-study corrections due to overconfidence in any study site's ability to meet the desired recruitment goal. A key challenge is the patchwork of EHR workflow processes that exist at CTSA institutions that can hamper efforts to standardize cohort discovery and uncover accurate data that is reflective of a site's potential participant pool. Logistical bottlenecks, regulatory and compliance approvals, limitations on data sharing, and lack of training and resources can all serve to frustrate or delay counts.[15]

Our model of partnering with CTSAs and leverage existing infrastructure to implement a cohort discovery process which remains flexible to account for variation in implementation among participating sites has been widely utilized and successful in terms of the site selection and expansion process. We rely on CTSAs having access to "big data" resources and tools for querying such as i2b2, ACT, and TriNetX. Our model does not require expensive new infrastructure to be built. Although this proposition works for CTSAs with research data warehouse capacity and processes designed for

rapid interrogation, it may not scale as easily to smaller institutions (eg, Federally Qualified Health Centers).

In addition, while numerous efforts are underway to develop better methods to share or standardize computable phenotype definitions or ontologies across research networks to support cohort discovery,[16–20] no consolidated repository currently exists. Rather than starting with a common data model and requesting that data teams populate a TIN data repository, we offer a tiered approach that respects the autonomy of individual CTSA hubs and their affiliates to fulfill the requirements as they are able. Although this model precludes perfectly harmonized answers across hubs, and sometimes prevents hubs from running the entire query, we have found it to be more equitable in enabling at least partial cohort assessment by all hubs. In addition to flexibility in running partial cohort numbers, we have found that successful requests include clear instructions and enough context to ensure hubs are interested in the study and not just running counts.

A limitation of our model is that it provides only counts, which are not necessarily predictive of recruitable participants. The counts themselves have limitations as well, such as being based on historical data that may not reflect the current state (eg, from a clinic that is no longer operational). A lack of domain knowledge about local data sources can be problematic. Also, data from EHRs are imperfect—records are complex, and can be inaccurate or incomplete.[21–23] Based on direct feedback from CTSA hubs as well as evaluating the burden on hubs using reported runtime data shown in Table 1, we have determined this resource to be most successful with broad queries rather than those that are surgically precise. With the intent of determining whether recruitment is feasible at a site based on how representative a certain disease condition is, there is little value in requesting counts using exact inclusion/exclusion criteria. These factors informed our creation of a model that values broad versus precise phenotyping, allowing approximate counts from many hubs over precise counts from fewer hubs.

A major caveat is that cohort counts alone are insufficient to evaluate the suitability of a particular institution for study recruitment. An array of confounding factors, including availability of a local PI, budget allotments for site start-up, the hub's capacity to recruit diverse populations, and the nuances of local IRB approval, all come into play, and are beyond the scope of this paper. Our model also cannot measure the quality of counts, nor does it account for myriad other factors that must be taken into consideration when choosing a trial site, including finding an appropriate and engaged local PI, sufficient budget and staffing, clinician buy-in for referrals,

Table 1. Detailed cohort assessment metrics and phenotype information

Query	Year	Disease Domains	Average runtime (h:m)	Response Rate (%)	Opt out rate (%)	No. of tiers in query request	No. of variable types	Dx codes	Px codes	Med	Encounter type	Lab	Temporal	Insurance type	Vital status
Query No. 1	2018	Hemo	3:30	67	6	1	3	X							X
Query No. 2	2018	Neuro	3:30	83	21	5	3	X		X					X
Query No. 3	2018	Cardio	2:0	76	10	3	3	X		X					X
Query No. 4	2018	Cardio	4:15	77	16	3	4	X	X		X		X	X	X
Query No. 5	2019	Cardio	3:15	70	22	2	3	X	X				X		X
Query No. 6	2019	Infectious disease	5:45	63	26	3	4	X	X	X	X		X		
Query No. 7	2019	Cardio	3:15	77	16	2	1	X							
Query No. 8	2019	Neuro	3:30	81	23	1	2	X	X						X
Query No. 9	2019	Cardio	2:0	72	32	1	2	X	X						
Query No. 10	2019	Audiology	4:15	74	40	3	2	X	X				X		
Query No. 11	2019	Cardio	3:30	66	33	3	2	X	X				X		X
Query No. 12	2019	Cardio	3:15	77	20	2	1		X				X		
Query No. 13	2019	Pediatric trauma	3:30	80	19	2	1	X							
Query No. 14	2019	Neuro	1:30	77	16	1	2	X							X
Query No. 15	2019	Neuro	5:0	71	33	2	3	X	X					X	X
Query No. 16	2019	Cardio	:45	67	36	1	3	X	X						X
Query No. 17	2019	Neuro	7:30	74	13	4	5	X	X				X	X	X
Query No. 18	2019	Neuro	4:0	80	15	5	2	X	X			X	X		
Query No. 19	2019	Behavior medicine	2:45	73	25	1	3	X	X				X		X
Query No. 20	2019	Neuro	2:15	74	10	2	1	X							
Query No. 21	2020	Neurology	2:15	79	21	3	2	X	X				X		
Query No. 22	2020	Pediatric	6:30	68	20	18	2	X				X			
Query No. 23	2020	Allergy	2:25	64	31	2	3	X		X	X		X		
Query No. 24	2020	Neuro	2:45	70	15	1	2	X	X						X
Query No. 25	2020	Infectious disease	2:45	91	17	2	3	X			X	X			
Query No. 26	2020	Vaccine	2:15	71	23	1	1	X							
Query No. 27	2020	Gastro	7:30	72	34	10	2	X	X				X		
Query No. 28	2020	Infectious disease	3:30	59	47	3	3	X			X	X	X		
Query No. 29	2020	Pediatric	3:30	52	29	3	2	X		X			X		
Query No. 30	2020	Neurology	6:30	80	28	2	2	X			X		X		

competing trials, and appeal of the study to potential participants. These issues are beyond the scope of this report but should be recognized.

Finally, the RIC methodology described here cannot be successful without establishing and maintaining trust with partnering institutions. Continued participation requires incorporating feedback about what is working and what is not. Communication is paramount. Participating sites want to be continuously informed, to learn how their counts and participation metrics compare with others. Also, given that the counts are only 1 aspect of site selection, potential sites need to be kept abreast of the other elements in the process and where their institution stands in terms of task completion. If their hub is not selected, reasons should be provided.

Based on our work with multiple TIN trials and diverse investigator teams, a completed EHR-based cohort assessment appears to be a strength for investigators in applications for funding. Although we do not yet have quantitative evidence to show that completing the query across the CTSA consortium has led to meeting anticipated study recruitment goals, the resource has been favorably highlighted in multiple summary statements and grant reviewer comments. Investigators who have worked with the RIC to complete the EHR-based cohort assessment have shared anecdotal reviewer feedback summarizing the value of the resource in terms of:

- Estimating the approximate recruitment population at existing or potential study sites,
- Obtaining feasibility numbers across the CTSA consortium,
- Identifying the appropriate number of study sites to include in the proposal, and
- Identifying backup sites that would be interested in participating if recruitment goals are not met.

The RIC resource line solution presented here is centered around helping hubs retrieve and return participant population estimates as part of a larger TIN process of engaging CTSA hubs and Affiliates in fulfilling Expression of Interest for specific trials. Our goal is to provide an efficient process to onboard new sites that demonstrate the interest in and ability to serve as a site in a multicenter trial. This work is generalizable and can be used by other programs considering a decentralized network-wide participant count request or reporting use case.

Future directions

Further evaluation will be conducted as sites begin to enroll participants into RCTs that received early EHR-based cohort assessment support. This information will help determine whether conducting pretrial cohort assessment to determine recruitment feasibility increases RCT enrollment. Future improvements to the EHR-based cohort assessment service line under consideration include improving methods for sharing queries directly from query platforms (eg, TriNetX and EPICs Slicer Dicer). As part of our work to understand capacity and utility of these self-service query platforms, the RIC conducted user testing and provided feedback based on our supported use cases.

CONCLUSION

Our iterative phenotype development process and socialization model for CTSA hubs have been successfully developed, deployed, and disseminated. Hubs have expressed appreciation for the simplicity and inclusiveness of our model and RCT PIs have expressed strong satisfaction in both the timeliness and utility of reported results. Our detailed description of the cohort assessment process can be replicated by other institutions desiring to support decentralized solutions in support of multicenter studies.

AUTHOR CONTRIBUTIONS

SN, BD, PH, and CHW led the design of the work. SN, BD, DH, RG, BL, TB, and CW contributed to the data acquisition or analysis. All authors contributed to result interpretation and manuscript writing and all authors approved the final draft for publication.

CONFLICT OF INTEREST STATEMENT

None declared.

DATA AVAILABILITY

The data underlying this article are available within the article.

REFERENCES

1. North CM, Dougan ML, Sacks CA. Improving clinical trial enrollment—in the Covid-19 era and beyond. *N Engl J Med* 2020; 383 (15): 1406–8.
2. Kost RG, Mervin-Blake S, Hallarn R, *et al.* Accrual and recruitment practices at clinical and translational science award (CTSA) institutions: a call for expectations, expertise, and evaluation. *Acad Med* 2014; 89 (8): 1180–9.
3. Cooley ME, Sarna L, Brown JK, *et al.* Challenges of recruitment and retention in multisite clinical research. *Cancer Nurs* 2003; 26 (5): 376–86.
4. Greene SM, Geiger AM. A review finds that multicenter studies face substantial challenges but strategies exist to achieve Institutional Review Board approval. *J Clin Epidemiol* 2006; 59 (8): 784–90.
5. Reis SE, Berglund L, Bernard GR, *et al.*; National Clinical and Translational Science Awards Consortium. Reengineering the National Clinical and Translational Research Enterprise: the strategic plan of the National Clinical and Translational Science Awards Consortium. *Acad Med* 2010; 85 (3): 463–9.
6. Bernard GR, Harris PA, Pulley JM, *et al.* A collaborative, academic approach to optimizing the national clinical research infrastructure: the first year of the Trial Innovation Network. *J Clin Transl Sci* 2018; 2 (4): 187–92.
7. Murphy S, Wilcox A. Mission and sustainability of informatics for integrating biology and the bedside (i2b2). *EGEMS (Wash DC)* 2014; 2 (2): 1074.
8. Dobbins NJ, Spital CH, Black RA, *et al.* Leaf: an open-source, model-agnostic, data-driven web application for cohort discovery and translational biomedical research. *J Am Med Inform Assoc* 2020; 27 (1): 109–18.

9. ATLAS—A unified interface for the OHDSI tools—OHDSI. https://www.ohdsi.org/atlas-a-unified-interface-for-the-ohdsi-tools/ Accessed July 13, 2021.

10. Topaloglu U, Palchuk MB. Using a federated network of real-world data to optimize clinical trials operations. *JCO Clin Cancer Inform* 2018; 2: 1.

11. Visweswaran S, Becich MJ, D'Itri VS, *et al*. Accrual to Clinical Trials (ACT): a Clinical and Translational Science Award Consortium Network. *JAMIA Open* 2018; 1 (2): 147–52.

12. Selby JV, Grossman C, Zirkle M, *et al*. Multistakeholder engagement in PCORnet, the National Patient-Centered Clinical Research Network. *Med Care* 2018; 56 (Suppl 1): S4–5.

13. Danciu I, Cowan JD, Basford M, *et al*. Secondary use of clinical data: the Vanderbilt approach. *J Biomed Inform* 2014; 52: 28–35.

14. Harris PA, Taylor R, Thielke R, *et al*. Research electronic data capture (REDCap)–a metadata-driven methodology and workflow process for providing translational research informatics support. *J Biomed Inform* 2009; 42 (2): 377–81.

15. Obeid JS, Beskow LM, Rape M, *et al*. A survey of practices for the use of electronic health records to support research recruitment. *J Clin Transl Sci* 2017; 1 (4): 246–52.

16. Richesson RL, Smerek MM, Blake Cameron C. A framework to support the sharing and reuse of computable phenotype definitions across health care delivery and clinical research applications. *EGEMS (Wash DC)* 2016; 4 (3): 1232.

17. Weng C, Tu SW, Sim I, *et al*. Formal representation of eligibility criteria: a literature review. *J Biomed Inform* 2010; 43 (3): 451–67.

18. Zhang H, He Z, He X, *et al*. Computable eligibility criteria through ontology-driven data access: a case study of hepatitis C virus trials. *AMIA Annu Symp Proc* 2018; 2018: 1601–10.

19. Anderson N, Abend A, Mandel A, *et al*. Implementation of a deidentified federated data network for population-based cohort discovery. *J Am Med Inform Assoc* 2012; 19 (e1): e60–7.

20. Verchinina L, Ferguson L, Flynn A, *et al*. Computable phenotypes: standardized ways to classify people using electronic health record data. *Perspect Health Inf Manag* 2018; (Fall): 1–8.

21. Hripcsak G, Albers DJ. Next-generation phenotyping of electronic health records. *J Am Med Inform Assoc* 2013; 20 (1): 117–21.

22. Weiskopf NG, Weng C. Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *J Am Med Inform Assoc* 2013; 20 (1): 144–51.

23. Weiskopf NG, Rusanov A, Weng C. Sick patients have more data: the non-random completeness of electronic health records. *AMIA Annu Symp Proc* 2013; 2013: 1472–7.