The design and implementation of an open-source, data-driven cohort recruitment system: the Duke Integrated Subject Cohort and Enrollment Research Network (DISCERN)

Jeffrey M Ferranti,1,2 William Gilbert,2 Jonathan McCall,3 Howard Shang,2 Tanya Barros,2 Monica M Horvath2

ABSTRACT

Objective Failure to reach research subject recruitment goals is a significant impediment to the success of many clinical trials. Implementation of health-information technology has allowed retrospective analysis of data for cohort identification and recruitment, but few institutions have also leveraged real-time streams to support such activities.

Design Duke Medicine has deployed a hybrid solution, the Duke Integrated Subject Cohort and Enrollment Research Network (DISCERN), that combines both retrospective warehouse data and clinical events contained in prospective Health Level 7 (HL7) messages to immediately alert study personnel of potential recruits as they become eligible.

Results DISCERN analyzes more than 500,000 messages daily in service of more than 12 projects. Users may receive results via email, text pages, or on-demand reports. Preliminary results suggest DISCERN’s unique ability to reason over both retrospective and real-time data increases study enrollment rates while reducing the time required to complete recruitment-related tasks. The authors have introduced a preconfigured DISCERN function as a self-service feature for users.

Limitations The DISCERN framework is adoptable primarily by organizations using both HL7 message streams and a data warehouse. More efficient recruitment may exacerbate competition for research subjects, and investigators uncomfortable with new technology may find themselves at a competitive disadvantage in recruitment.

Conclusion DISCERN’s hybrid framework for identifying real-time clinical events housed in HL7 messages complements the traditional approach of using retrospective warehoused data. DISCERN is helpful in instances when the required clinical data may not be loaded into the warehouse and thus must be captured contemporaneously during patient care. Use of an open-source tool supports generalizability to other institutions at minimal cost.

INTRODUCTION

The capacity to identify, evaluate, and recruit potential research participants is crucial to the success of clinical and health services research. Many institutions, however, have difficulty finding potential subjects1,2 and thus may fail to meet accrual goals or deadlines. A continuing emphasis on harnessing clinical and administrative data sources for multiple uses, coupled with increasingly sophisticated health information technology (HIT) and informatics capabilities, may substantially improve the efficacy and cost-effectiveness of clinical trials recruitment efforts.3–6

Our institution, Duke Medicine, comprises two community hospitals, an academic facility (Duke University Hospital (DUH)), and more than 130 affiliated outpatient clinics. We report on the design and initial implementation of the Duke Integrated Subject Cohort and Enrollment Research Network (DISCERN), which combines data from the enterprise data warehouse and prospective Health Level 7 (HL7) messages to enable real-time alerting of potential enrollees for clinical, quality improvement, and research staff. We discuss several use cases, describe a preconfigured, self-service DISCERN task through our researcher portal, and compare DISCERN to existing real-time recruitment methods.

BACKGROUND

Well-conducted, adequately powered clinical trials are essential to evidence-based medicine and quality-improvement activities. The success of such trials, however, depends largely on the capacity to identify, recruit, and enrol sufficient numbers of appropriate subjects within a highly constrained time frame.7 Recruitment efforts represent a substantial proportion of study resources and may consume as much as 50% of the total cycle time for a clinical trial.8 Findings from studies by Haidich and Ioannidis9 suggest that clinical trials are more likely to reach accrual targets when they achieve high rates of enrollment within their first 2 months. Conversely, failing to reach recruitment goals is implicated in the lack of success of numerous studies10 and imposes significant burdens on patients and healthcare consumers by delaying the translation of new therapies from the laboratory to the bedside.11

In some cases, recruitment targets are missed owing to inaccurate estimates of the pool of eligible participants.12–14 Many health systems possess extensive and rapidly growing data repositories that, when appropriately integrated and made accessible to clinical personnel, provide rich sources of information on demographics, clinical visits, diagnoses, procedures, and clinical orders that may be evaluated to identify individuals of interest to researchers. Various academic research centers have
Therefore sought to leverage their data resources to identify potential subjects. Commercially available and ‘homegrown’ applications have both attempted to leverage electronic health records (EHRs) for cohort identification, with varying levels of success and generalizability.

At Duke Medicine, we similarly sought to improve cohort discovery by creating Duke Enterprise Data Unified Content Explorer (DEDUCE), a Web-based query tool that affords the Duke research community access to our enterprise data warehouse, the Decision Support Repository (DSR), which integrates and stores more than 12 years of clinical and billing data from across the health system. Researchers who have been trained and granted appropriate access can query the DSR according to study criteria and create reports summarizing the number of eligible potential subjects, diagnoses, and demographics, as well as clinical locations visited.

However, we soon realized that a tool using warehoused data would not suffice for three key scenarios. First, it is possible that the associated databases do not capture the necessary information needed to support a study. For example, at DUH, the timestamp noting a nurse’s view and acknowledgment of a physician order is not loaded into the DSR—an omission that creates a barrier to study designs examining the timing associated with STAT orders.

Second, eligibility may depend upon time-sensitive criteria. Such specifications may include not just the wide variety of clinical events communicated across HIT systems (such as emergent diagnoses, urgent physician orders, lab results, transfers between care settings, or presentation to an emergency department) but also the timing of these events relative to one another.

Third, investigators who have identified a cohort may still require options for outpatient monitoring after treatment has been provided. Depending on the subject area, a data warehouse integrates information from frontline clinical systems after a latency period ranging from 1 to 50 days, because the context needed to place data points into the schema often must be reconciled from numerous frontline systems serving distinct clinical workflows. In many instances, data fields needed for warehouse integration are only complete within each of these source systems at patient discharge. This precludes using DSR information to identify patients in real time as their clinical condition or location changes.

Such temporal challenges are not new to healthcare organizations. For some time, hospitals, medical centers, and health systems have relied on automated alerting and decision-support systems to serve their clinical and research missions. In a seminal configuration (arrangement of the necessary retrospective and prospective queries), and (3) reporting (provision of alert).

Stage 1: Definition of the DISCERN job

Because HL7 is an interoperability standard and not readily digestible by a casual clinical user, the first stage necessarily comprises a consultation between the DISCERN developer and the researcher to define functional requirements. The developer has an understanding of data provenance and is thus best suited to determine how HL7 messages may meet the user’s needs. Users whose needs can be accommodated by a self-service retrospective query of the DSR may be directed to the DEDUCE portal. To use DISCERN, all users must have an active, institutional review board (IRB)-approved protocol explicitly stating that DISCERN is part of the recruitment strategy. DISCERN
Stage 2: Configuration of the DISCERN job

To reason over HL7 messages passed through the integration broker, we selected an open-source, modular tool that uses a widely available programming language. Mirth software (http://www.mirthcorp.com/ (WebReach)) is an open-source, standards-based solution for healthcare system interoperability that receives, interprets, processes, and sends healthcare information (such as HL7 messages) over a variety of protocols to support the exchange of health information. Mirth is packaged with a Java-based dashboard tool (Mirth Connect Administrator) that provides a graphical user interface (GUI). Within this dashboard, users may supply JavaScript to process HL7 messages exchanged by the integration broker.

A Mirth instance comprises a series of channels configurable within Mirth Connect Administrator. A channel is an interface that consists of a connector (ie, data source specification), one or more filters (instructions to sort incoming data and accept or reject), one or more transformers (instructions to modify accepted data), and one or more destination connectors (transmission of transformed data externally or to another channel).

To create a channel, the developer uses the Mirth Connect Administrator dashboard to perform three general steps (each of which is a distinct tab within the dashboard): (1) configuration of information acceptance, (2) configuration of information transmission, and (3) application of any instructions using a short, channel-associated JavaScript (eg, apply filters; call other code). In this context, ‘information’ may include HL7 messages, SQL queries, or other formats accessible through an ODBC interface. Mirth can also load a file of ‘test’ HL7 messages to be passed through the channel architecture in order to assess the fidelity of query processing for a project’s objective. Outbound information can be sent to other channels or written to a database.

A DISCERN project can use multiple channels to accomplish objectives; contrariwise, a single channel can serve multiple projects. An unlimited number of channels can be created in Mirth, allowing developers to serve scenarios wherein required data points arrive via multiple HL7 messages. Mirth also permits storage of data in temporary tables or memory so that they can be reasoned over using additional information—either by way of HL7 messages or a separate query of another database (eg, a research subject registry). Figure 1 provides an overview of how DISCERN uses Mirth to process HL7 messages.

To further illustrate the flexibility of Mirth channel architecture, we present a specific case from the Duke Heart Center in which a group of physicians desires immediate notification if one of their cardiac surgery patients is readmitted to the hospital or emergency department following discharge for the original surgery. Online appendix 1 describes the technical details for configuring this scenario, which has both prospective and retrospective query steps. The first Mirth channel is configured to examine and route all inbound HL7 messages to the integration broker. In this case, when an admission-discharge-transfer (ADT) message is intercepted, this first channel recognizes the message type and acts as ‘traffic control’ to route that information to a second channel dedicated to processing ADT messages. (This channel has a variety of instructions relevant to any project where ADT data are required for a real-time query, not just the example at hand.)

This second channel accepts the incoming data and composes with JavaScript instructions bound to the channel, which in this case are to send an SQL query to the DSR asking whether the patient represented in the HL7 message is present in a relational database that contains all recently discharged Duke Medicine cardiac surgery patients. If the ADT message pertains to a patient in the database, it returns a value of ‘True.’ This ADT channel is configured so that receiving a value of ‘True’ instigates the sending of information to a third channel responsible for real-time notification.

Channel 3 is a generic channel built to create an email alert for DISCERN users. Channel 2 sends Channel 3 the name of the provider to be notified, the identifiers of the readmitted patient, and the patient’s current location. An email with these data is sent to alert the physicians that a recent cardiac surgical patient has been readmitted to a Duke Medicine hospital or emergency department.

This case demonstrates the unique hybrid query scenario for a project that would require both real-time and archival data. At Duke Medicine, an approach that was restricted to a retrospective examination of the data warehouse would be capable of alerting personnel to a readmitted patient only well after the care event had occurred. In many cases, this would take place after a patient is discharged, and the billing process has...
begun—too late to ensure that the providers responsible for the prior surgery are made aware of the patient’s readmission and treatment.

This example showcases just one portion of the channel architecture that enables numerous DISCERN projects. Channels can support multiple queries, as JavaScript details can apply sorting logic and other filters to HL7 messages to suit a given project. When a new project is brought to DISCERN developers, it is integrated into the existing channel architecture, because the developer ultimately seeks to minimize the number of active channels in order to easily manage distinct projects. The developer may also consolidate channels with a generic functionality shared by multiple projects, such as sending a text page or parsing ADT data. For any DISCERN project, the two final channels in the DISCERN workflow will (1) notify the researcher of any identified patients meeting study criteria and (2) place the information into a DISCERN database with its own subschema within the DSR for the purpose of later error-trapping if needed. The DISCERN service runs automatically without further interaction until deliberately stopped.

### Stage 3: DISCERN reporting
In the final stage of the DISCERN workflow, a notification may be sent to the researcher immediately (eg, SMS message, text page, or email) or in aggregate (spreadsheet file or on-demand report written in a business intelligence program). When designing requirements in the initial consultation, the developer would advise the end user as to which data-reporting method is likely to yield the best results and not subject the user to alert fatigue. DISCERN jobs are automatically triggered to cease reporting once an IRB protocol expires.

### RESULTS
DISCERN currently reads 500,000 messages per day in service of 12 distinct projects organized into 11 Mirth channels. A summary of current DISCERN use cases is shown in table 1. While most DISCERN jobs concern cohort recruitment, others have applied similar principles to quality-improvement projects, thus demonstrating model’s versatility. The reporting method correlates with the window of time available to the researcher to recruit the subject, given the study parameters. The capability of

<table>
<thead>
<tr>
<th>Type</th>
<th>Delivery</th>
<th>Use case</th>
<th>Filter criteria</th>
<th>Data points provided</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research/ recruiting</td>
<td>Email and text page</td>
<td>Identify patients with positive blood bacteria cultures currently admitted to DUH</td>
<td>Lab result</td>
<td>Patient identifiers; lab result, inpatient location</td>
</tr>
<tr>
<td>Research/ recruiting</td>
<td>Text page</td>
<td>Identify DUH newborns at risk of ischemic brain injury to collect umbilical cord blood cells immediately after birth</td>
<td>Location; order entry for ‘total body cooling’</td>
<td>Patient identifiers; inpatient location</td>
</tr>
<tr>
<td>Research/ recruiting</td>
<td>Email</td>
<td>Identify patients admitted to two DUH pediatric locations or seen in the emergency department by an attending oncologist, with the goal of detecting possible dangerous fungal infections.</td>
<td>Location, admitting physician</td>
<td>Patient identifiers, admitting physician, diagnosis, inpatient location</td>
</tr>
<tr>
<td>Research/ recruiting</td>
<td>On-demand reporting</td>
<td>Identify patients with new ICD-9 diagnoses for atrial fibrillation or atrial flutter for recruitment into an anticoagulant study</td>
<td>ICD-9 code</td>
<td>Patient identifiers, upcoming appointments</td>
</tr>
<tr>
<td>Research/ recruiting</td>
<td>On-demand reporting</td>
<td>Identify chronic gout sufferers with high uric acid lab results</td>
<td>ICD-9 codes, age, lab value</td>
<td>Patient identifiers, demographics, lab result, upcoming appointments</td>
</tr>
<tr>
<td>Research/ recruiting</td>
<td>On-demand reporting</td>
<td>Identify adult asthma sufferers</td>
<td>Age and ICD-9 code</td>
<td>Patient identifiers, upcoming appointments</td>
</tr>
<tr>
<td>Research/ recruiting</td>
<td>On-demand reporting</td>
<td>Recruit adolescent girls who have received one or two doses of the HPV vaccine series</td>
<td>Age and ICD-9 code</td>
<td>Patient identifiers, upcoming appointments</td>
</tr>
<tr>
<td>Research/ monitoring</td>
<td>Text page</td>
<td>From a list of patients registered for a phase II pediatric clinical trial receiving insulitol at DUH, identify instances when creatine values rise above the threshold value mandating discontinuation of investigational drug</td>
<td>List of patients enrolled in IRB protocol; lab result</td>
<td>Patient identifiers, inpatient location, lab result</td>
</tr>
<tr>
<td>Quality improvement</td>
<td>Pop-up alert sent to nearest computer in nursing unit</td>
<td>Immediately alert neonatal intensive care nurses when an urgent ‘STAT’ order is entered by a physician in the computerized provider order entry system by sending a pop-up alert to the computer closest to the affected patient</td>
<td>Computerized provider order entry order type; patient location</td>
<td>Patient identifiers</td>
</tr>
<tr>
<td>Quality improvement</td>
<td>On-demand reporting</td>
<td>Identification of high-risk pediatric patients who should be prioritized for H1N1 vaccine delivery during a vaccine shortage</td>
<td>Age and ICD-9 code</td>
<td>Patient identifiers; upcoming appointments</td>
</tr>
<tr>
<td>Quality improvement</td>
<td>Email</td>
<td>Identification of pediatric patients with an admitting diagnosis indicating asthma</td>
<td>Age; text matching for asthma-related terms</td>
<td>Patient identifiers; inpatient location; attending physician</td>
</tr>
<tr>
<td>Quality improvement</td>
<td>Email</td>
<td>Identify Duke Medicine Heart Center surgery patients recently readmitted at a Duke Medicine hospital and alert the discharging physician</td>
<td>Patient discharged by Duke Heart Center within last 31 days</td>
<td>Patient identifiers; inpatient location</td>
</tr>
</tbody>
</table>

DUH, Duke University Hospital; ICD-9, International Classification of Diseases, 9th revision.
the Mirth engine to listen prospectively to real-time HL7 messages is central to the success of the DISCERN workflow. For any use case in table 1 that must report subject identification immediately (ie, by text page or email), a method that places this information into databases or data marts for later query could not have instantly notified the study coordinators of a potential recruit.

**DISCERN use cases for patient recruitment**

**Umbilical-cord blood collection**

DISCERN was used in a study involving the prospective collection and intravenous readministration of umbilical cord blood cells, which is hypothesized to be a neuroprotective adjunct to total-body cooling in cases of neonatal hypoxic ischemic encephalopathy (HIE). The study protocol mandates that cord blood be procured and prepared immediately after birth and administered within the first days of life. Obstetricians were asked to collect cord blood at every high-risk delivery; however, given the acuity of these deliveries, there were multiple missed opportunities for recruitment, particularly among infants born late at night.

From July through December of 2009, six infants delivered at DUH underwent total body cooling, but the study team was not notified of candidates for 6–24 h. As a result, no cord blood units were procured, and no patients were enrolled in the trial. A DISCERN process was constructed that (1) looked for a nursing order of ‘total body cooling,’ (2) paged a study coordinator to start the consent process, and (3) paged the study investigator, who could contact the obstetrician to ensure cell procurement. By backtracking to deliveries of children with hypothermia orders but no cord-cell collections, we identified issues with the stocking of the sterile cell procurement kits as well as with obstetrician availability, underscoring not only DISCERN’s usefulness for patient recruitment, but also its application to process and quality improvement. Once the DISCERN process was in place and the cord-blood collection process was optimized, nine patients were cooled, of whom eight were enrolled in the study from January to November of 2010.

**Human papillomavirus vaccine study**

DISCERN was used to enrol patients in a study investigating the timing of administration of a human papillomavirus (HPV) vaccine series; specifically, the effects on the immune system if doses are not given at the specified intervals (0, 2, and 6 months). The study requires blood to be drawn from subjects at the administration of the third vaccine dose and again 1 month later. Researchers thus needed to be able to quickly find patients residing in the appropriate window of the vaccine series.

To identify a cohort for this study, DISCERN scanned for adolescent girls who received the first or second shot in the series, as identified by Current Procedural Terminology (CPT) billing codes. Results were made accessible to researchers through on-demand reporting that showed patient identifiers and upcoming appointments. The study team then contacted the attending physician to obtain permission to approach the patient at the next appointment.

In the first 10 months of the study, subjects were recruited using traditional pathways, which resulted in 43 enrollees out of 448 patients approached (an enrollment rate of 9.6%; figure 2). Starting in April of 2010, DISCERN was used to identify potential recruits. Over 4 months, there were 62 enrollees out of 421 patients approached—an enrollment rate of 14.7%. Not only was this 55% higher (Pearson χ² test, p=0.02), but the recruitment rate increased from an average of 4.3 subjects/month to 15.5 subjects/month. The study team also reported a reduction in eligibility screening time, from >10 h/week to 4 h/week.

**Pediatric asthma study**

As part of the quality-improvement project, a group of DUH pediatricians sought to identify all pediatric asthma patients admitted to the hospital. One challenge presented by this project was that children with new diagnoses of asthma at DUH would not have an indication of their condition in the DSR. DISCERN was used to monitor multiple HL7 message types identifying patients who were (1) <20 years old, (2) admitted to a general pediatrics floor, and (3) received an admitting diagnoses of asthma as entered into the computerized provider order entry system. Because admitting diagnoses are captured only as free text by computerized provider order entry, a DISCERN channel was configured to use regular expression matching to search the text string for series of terms indicative of asthma, as well as common abbreviations and misspellings. Because a diagnosis of asthma is difficult to make in young children, the search term

![Figure 2](https://academic.oup.com/jamia/article-abstract/19/1/e68-e75/728696)
list was expanded to include more general terms such as ‘wheezing.’ When the triggering conditions for the channel were met, DISCERN sent an email that contained the admitting physician, matching admission diagnosis, and patient medical record number to four other pediatricians. In a 1-month pilot phase, DISCERN identified 15 patients, 13 of whom were true asthmatics included in the QI project. Given the architecture of HIT systems at DUH, this work would not have been possible using a researcher query tool such as i2b2, as the admitting diagnosis is not available for database query until the medical billing process has been completed postdischarge. Without DISCERN, physicians would have been obliged to resort to daily manual checking of EHRs for eligible patients.

**Self-service DISCERN applications**

Due to the need to maintain data provenance, clinicians wishing to use the DISCERN tool for a unique query must work with an analyst who configures Mirth. However, upon fulfilling several requests, it became clear that a simple and common DISCERN use case could be automated within the context of our DEDUCE researcher portal. One of the most frequent requests was for access to timely information regarding future clinic appointments for a list of patients. Within the DEDUCE portal, researchers may (in a self-service manner) either define a cohort by querying the organizational data warehouse or upload a list of medical record numbers. Within this portal, they now have access to a DISCERN ‘button’ that with one click provides a current report detailing all upcoming clinic appointments, including date, time, clinic name, provider, clinical service, visit type, and visit reason. This allows trial coordinators to meet the potential recruit’s physician at the next healthcare encounter and request permission to offer a trial to the patient. In this way, DEDUCE developers are entirely removed from the logistics of defining a new DISCERN query because the underlying Mirth channels already exist. As additional reusable use cases are identified, they will be tested and automated for self-service within the DEDUCE researcher portal.

**DISCUSSION**

**Summary**

Clinical investigations and quality-improvement research often require real-time data on patient care actions. Novel HIT applications that offer access to retrospective data in support of clinical trial research such as STRIDE, i2b2, and DEDUCE have been described in the literature. DISCERN builds upon the capabilities afforded by these applications by providing a hybrid framework for clinical trial recruitment that uses HL7 messages containing real-time data on clinical events to complement more traditional approaches utilizing warehoused data. The functionalities afforded by access to real-time data become particularly important when timing issues are critical to study enrollment, multiple events within distinct HIT systems must be cross-evaluated to confirm eligibility, and/or desired data are not permanently captured by the target clinical application for warehouse loading. We also posit that in addition to improving the mechanics of data capture, DISCERN’s filtering capabilities can contribute to better clinical research associate workflow by expediting in-person prescreening for trial eligibility during clinic visits.

**Comparison to other subject recruitment tools**

Most reported cohort recruitment tools rely on retrospective information and are not able to use data collected prospectively during clinical care. Whereas warehoused data are often standardized and organized into cross-referenced tables, clinical data from operational systems are newly created, quickly communicated, and often standardized only to the extent required for intercommunication. Although warehoused data may have timely refresh rates, these are typically on the order of daily or weekly. Thus, different tools are needed to tap into each data type, particularly when the time window for recruiting a subject is narrow or study eligibility criteria are temporally constrained.

The use of HL7 messaging specifically to support clinical trial recruitment is not widespread. In a recent conference proceeding, Weber and colleagues describe two pilot studies in which the Java-based Esper engine was used to deploy custom Java classes that could read HL7 messages and filter them according to multiple recruitment criteria using a variation of SQL. As with Mirth, the Esper engine is open source, but the authors note that an adapter for HL7 messages had to be manually coded, suggesting that significant local configuration was required. In addition, the Esper-based system is purely real-time in nature and presently does not permit consulting with relational databases, although such hybrid capabilities are slated for future development. In contrast, Mirth is purpose-designed as a tool for healthcare interoperability and has built-in HL7-reading functionality within a GUI; in addition, its hybrid design facilitates the use of both real-time data and retrospective, warehoused information.

Most other cohort recruitment tools that rely on real-time data sources typically incorporate scripted searches executed at repeated, frequent intervals. For example, the RealTime Recruiting tool scans emergency department registration records every 2 min for criteria that match study requirements, and Cardozo and colleagues report on a program that searches emergency department registration records at 15 min intervals. These methods, however, are limited to the body of information contained within a single system, hindering their generalizability to other settings. In a recent paper, Nelson and colleagues report on a near–real-time alerting system that was deployed at the Mayo Clinic to monitor for instances of septic shock. A research data mart was queried hourly for a complex series of predefined rules and was updated with information from ICU patients’ EHRs as frequently as every 15 min, thus receiving information from numerous systems. Such a system offers the benefit of being relatively simple to implement; however, the continuous query model may not scale well in larger populations. DISCERN, on the other hand, draws from a confluence of health-system-wide data streams, as opposed to a single care setting or department. Instead of performing constant queries, the Mirth engine filters HL7 messages for applicability to an existing DISCERN job.

In addition, a number of models described in the literature target only a single point of contact in a specific role (eg, a care provider or principle investigator). Data entered during patient evaluations and treatment planning are screened in real time for criteria that, if met, trigger a pop-up prompt to alert the provider of eligibility during the examination. However, Grundmeier and colleagues report that this method is not always easily integrated with provider workflow. They note that physicians often ignore pop-up alerts, citing lack of sufficient time both to treat the patient and explain a study for recruitment purposes.

In contrast, DISCERN allows multiple persons performing a variety of roles to be notified following a trigger, and the mode and frequency of an alert can be tailored to the project’s needs. Further, the various permutations of staff and types of
notification can be used synchronously. Investigators and staff can opt to have DISCERN batch reports every few hours or send a page immediately upon encountering a trigger event. This type of customization, useful in improving recruitment for a particular study, also lends itself to a variety of investigatory avenues.

Given that real-time monitoring of clinical and laboratory messaging and real-time data warehousing have been proposed and implemented for biosurveillance purposes, such platforms could be leveraged to create notification of potential recruits as a side benefit of real-time extract-transform-load processing. In such a scenario, clinical data feeds would be loaded in real time to give researchers full access to all data points when developing a recruitment query. This ‘all or nothing’ approach would be highly problematic for many organizations, because the cost of real-time warehousing increases sharply with the amount of data incorporated and the frequency of data refresh rates. In contrast, the DISCERN model allows us to focus real-time data capture only where it is needed, using a free, open-source application. DISCERN thus offers a tractable and scalable form of real-time data warehousing.

Limitations and lessons learned
The DISCERN framework is easily configurable and most extensible to organizations having both an enterprise feed of HL7 messages as well as a clinical data warehouse. However, it is not a stand-alone tool but a service provided by a technical team. At our organization, the initial consultation between the DISCERN developer and the end user is critical both for ensuring success and for verifying that there is a legitimate need for real-time data that exceeds the capabilities of a self-service researcher query tool such as our DEDUCE platform or i2b2.

As the service increases in popularity, issues of scalability may present challenges, and our organization may need to hire more staff to translate and deploy customer requests. At present, however, our single DISCERN developer spends only 8 h/week in service of 12 projects, whereas the six analysts providing support for 50 projects work 40 h/week. This enhanced efficiency achieved by DISCERN is due to several factors. Once a channel is built, it is easily reused for other projects and configurable in the drag-and-drop Mirth interface. Any associated JavaScript is easily modifiable to suit a new yet familiar query request. Once a DISCERN query is released, it persists without maintenance until explicitly deactivated. Finally, as described above, a preconfigured, commonly requested DISCERN task has been added for self-service within the DEDUCE portal. For reasons related to data-provenance concerns, we believe it inadvisable to give researchers unlimited access to a DISCERN-like tool; however, we expect to add more self-service tasks in the future to complement the ad hoc DISCERN service.

The DISCERN workflow is a revolutionary but disruptive technology that fundamentally changes the ground rules for subject recruitment. Because of efficiencies potentially afforded by DISCERN, its use can exacerbate competition for patients. As patients generally are not enrolled in multiple trials, negotiating the governance of DISCERN across multiple studies and clinics presents ongoing challenges, and we are presently establishing a governance structure to address these important issues. If clinicians dislike competing for subjects to enrol in their trials, they may reject the application—even if it significantly enhances recruitment. The cultural and anthropologic issues surrounding tools such as DISCERN deserve independent study and are not meant to be addressed by the present work. However, we note that no technology, including DISCERN, can address other key obstacles to recruitment—lack of patient interest in the trial, or a primary care doctor that is uncomfortable referring the patient.

A final major limitation of this work is that it is too early to judge definitively whether DISCERN has meaningfully increased the efficiency of clinical trial recruitment across Duke Medicine. The studies using DISCERN are ongoing, and the perceived efficacy of the approach among our clinical research partners has engendered a reluctance to recruit patients without using the technology. We plan to conduct a randomized controlled trial in which one study coordinator group uses DISCERN, and another uses traditional methods, but have yet to find a clinical research team willing to reduce their DISCERN usage to allow for this comparison. In addition, deriving sensitivity and specificity statistics would require capture of false negatives (eg, real ‘hits’ missed by DISCERN) as well as true negatives (eg, the number of all potential hits correctly rejected by DISCERN). We are evaluating a number of DISCERN projects for the potential to conduct simultaneous manual chart review to collect false negatives, but only those query criteria easily identified through the EHR worklist can be evaluated using this method. Identification of true negatives, meanwhile, would require hardware sufficient to collect all HL7 messages and store them for later parsing by a scripting language.

Future work
As noted above, in cases where many different DISCERN requests follow the same pattern or recipe, developers will incorporate a DISCERN feature into the DEDUCE research portal. This integration of DISCERN into DEDUCE infrastructure allows centralized management of both retrospective and prospective data, potentially yielding economies of scale and freeing study investigators from burdensome data-management tasks. If this proves successful, we may extend DISCERN to the external research community by integrating it with the i2b2 heterogeneous structure currently in place at Duke Medicine that shares data with other National Institute of Health Clinical and Translational Science Award holders. In this way, interinstitutional i2b2 cohorts could be refined using real-time information.

Eventually, we plan to add functionality that would allow targeting of relevant clinical trial opportunities to patients’ health portal accounts, which provide Web access to lab results, vital measurements and medical history, upcoming appointments, online bill payment, and appointment scheduling.

CONCLUSIONS
Complementary approaches are needed to harness both retrospective and real-time data to identify potential study subjects and alert appropriate staff. The DISCERN framework serves a variety of clinical trial recruitment scenarios, including problematic variations in which study eligibility hinges on a confluence of specifications, clinical events, and temporal dependencies. In contrast to other cohort recruitment tools, the DISCERN model offers a highly flexible solution for prospective recruitment, with architecture extensible to any other organization that relies on HL7 messaging among clinical data systems. Our experience may help others seeking to deploy similar systems achieve success more quickly. More study is needed to definitively evaluate whether DISCERN goes beyond patient identification to measurably increase the speed of successful recruitment into clinical trials.
Acknowledgments The authors thank the Duke University Health System Information Management Group, R Calif, R Goldberg, D Tanaka, J Eckstrand, M Cotton, W Steinbach, J Wynn, B Vickery and C Walters for design guidance and the sharing of DISCERN use cases. We also thank E Hammond for critical review of the manuscript.

Funding This publication was made possible by Grant Number UL1 RR024128 from the National Center for Research Resources (NCRR), a component of the National Institutes of Health (NIH), and NIH Roadmap for Medical Research. Its contents are solely the responsibility of the authors and do not necessarily represent the official view of NCRR or NIH. Information on NCRR is available at http://www.ncrr.nih.gov/. Information on Re-engineering the Clinical Research Enterprise can be obtained from http://nrroadmap.nih.gov/cliclncalresearch/overview-translational.asp.

Competing interests None.

Provenance and peer review Not commissioned; externally peer reviewed.