Clinical Research Informatics: Challenges, Opportunities and Definition for an Emerging Domain

PETER J. EMBI, MD, MS, PHILIP R.O. PAYNE, PhD

Abstract

Objectives: Clinical Research Informatics, an emerging sub-domain of Biomedical Informatics, is currently not well defined. A formal description of CRI including major challenges and opportunities is needed to direct progress in the field.

Design: Given the early stage of CRI knowledge and activity, we engaged in a series of qualitative studies with key stakeholders and opinion leaders to determine the range of challenges and opportunities facing CRI. These phases employed complimentary methods to triangulate upon our findings.

Measurements: Study phases included: 1) a group interview with key stakeholders, 2) an email follow-up survey with a larger group of self-identified CRI professionals, and 3) validation of our results via electronic peer-debriefing and member-checking with a group of CRI-related opinion leaders. Data were collected, transcribed, and organized for formal, independent content analyses by experienced qualitative investigators, followed by an iterative process to identify emergent categorizations and thematic descriptions of the data.

Results: We identified a range of challenges and opportunities facing the CRI domain. These included 13 distinct themes spanning academic, practical, and organizational aspects of CRI. These findings also informed the development of a formal definition of CRI and supported further representations that illustrate areas of emphasis critical to advancing the domain.

Conclusions: CRI has emerged as a distinct discipline that faces multiple challenges and opportunities. The findings presented summarize those challenges and opportunities and provide a framework that should help inform next steps to advance this important new discipline.


Introduction

Clinical research is critical to the advancement of medical science and public health. Conducting such research is a complex, resource intensive endeavor comprised of a multitude of actors, workflows, processes, and information resources. Ongoing large-scale efforts have explicitly focused on increasing the clinical research capacity of the biomedical sector and have served to increase attention on clinical research and related biomedical informatics activities throughout the governmental, academic, and private sectors.1–8 Such programs and initiatives have served as significant catalysts for the emergence of a new sub-discipline of biomedical informatics focused on clinical research referred to as Clinical Research Informatics (CRI). The CRI space is growing rapidly and has already enabled significant improvements in the quality and efficiency of clinical re-

Affiliations of the authors: Department of Medicine and Center for Health Informatics, University of Cincinnati (PJE), Cincinnati, OH; Department of Biomedical Informatics and Center for Clinical and Translational Science, The Ohio State University (PROP), Columbus, OH. Both authors contributed equally to the preparation of this manuscript. The authors acknowledge the contributions of those who participated in our face-to-face session at the 2006 AMIA annual symposium and of members of the AMIA CRI working group who participated in phases of this research. In particular, the authors thank the following individuals listed alphabetically for their additional comments and other contributions to aspects of the preparation of this manuscript: Barbara Alving, MD; Suzanne Bakken, DNsC, RN; Charles Barr, MD, MPH; Tara Borlawsky, MA; Amar Chahal, MD, MBA; Christopher Chute, MD, Dr.P.H.; Milton Corn, MD; Don Detmer, MD; Bill Hersh, MD; Charles Jaffe, MD, PhD; Stephen Johnson, PhD; Srinivasc Kalluri; Stan Kaufman, MD; Rebecca Kush, PhD; Judith Logan, MD, MS; Daniel R. Masys, MD; Shawn Murphy, MD, PhD; Ricardo Pietroban, MD, PhD, MBA; and Ida Sim, MD, PhD; Justin Starren, MD, PhD.

Preliminary analysis and findings of phase one of this study were presented at the AMIA 2007 Annual Symposium and published in the AMIA 2007 Annual Symposium Proceedings.

Dr. Embi’s efforts in this research were supported in part by grants from the NIH/NLM (K22-LM008534, R01-LM009533). Dr. Payne’s efforts in this research were supported in part by grants from the NIH/NCI (P01-CA081534, R01CA134232) and NIH/NCCR (U54-RR024384).

Correspondence: Peter J. Embi, MD, MS, Center for Health Informatics, University of Cincinnati Academic Health Center, 231 Albert Sabin Way, PO Box 670840, Cincinnati, OH, 45267-0840; e-mail: <peter.embi@uc.edu>.

Received for review: 09/17/08; accepted for publication 02/12/09.
As CRI emerges as a highly valued area of activity, it is imperative that those working in and concerned with advancements in this space share an understanding of its scope and the range of challenges and opportunities facing the domain.

Background

Clinical Research

The United States NIH defines clinical research as:

“The range of studies and trials in human subjects that fall into the three subcategories:

1. Patient-oriented research. Research conducted with human subjects (or on material of human origin such as tissues, specimens and cognitive phenomena) for which an investigator (or colleague) directly interacts with human subjects. Patient-oriented research includes: (a) mechanisms of human disease, (b) therapeutic interventions, (c) clinical trials, or (d) development of new technologies.
2. Epidemiologic and behavioral studies.
3. Outcomes research and health services research.”

Several studies have illustrated that a lack of sufficient IT infrastructure and tools, as well as a reliance on workflows defined by historical precedence rather than optimal operational strategies, account for significant impediments to the expedient, effective, and resource-efficient conduct of clinical research activities. The rapid pace of biomedical science and the need for advances in medicine demand that the conduct of clinical research be timely, efficient, and yield high quality results. As a result, the importance of making clinical care data available for the secondary use in support of clinical research has become a competitive requirement for clinical and research enterprises. Moreover, the increasing complexity of clinical research and the challenges of regulatory requirements associated with conducting clinical studies have led to further changes in the clinical research landscape, including a trend towards conducting clinical trials in community practice settings, as opposed to the historical norm of conducting such studies in large Academic Health Centers (AHCs). The rapidly evolving and expansive clinical research landscape has led stakeholders in the clinical research environment to acknowledge and call for system-level solutions.

Consequently, clinical research is itself a domain in transition. Clinical researchers are faced with significant and increasingly complex workflow and information management challenges. The domain is also increasingly in the forefront of attention for the governmental, academic, and private sectors, all of whom have significant scientific and financial interests in the conduct and outcomes of clinical research efforts. Due to the preceding characteristics of the clinical research environment and the recognition that effective and efficient information access is critical to any solution to the many challenges faced by the domain, there has been a corresponding and rapid evolution of the biomedical informatics methods and tools specifically designed to address clinical research information management requirements.

Clinical Research Informatics

The evolution described above has led to efforts focused at the intersection of biomedical informatics and clinical research and the emergence of a domain that has become known as CRI. In addition to the reasons already noted, part of this evolution of CRI can be attributed to the extraordinary increase in the scope and pace of clinical and translational science advancements that have been catalyzed of late by the emergence of major funding programs such as those of the National Institute of Health’s (NIH) Road map initiative. A major goal of the NIH Road map involves programs to fundamentally re-engineer the way in which organizations translate basic science discoveries into practice.

Beyond the CTSA program, there are several other important efforts that are representative of CRI domain activities. Examples include 1) the NCI’s Cancer Biomedical Informatics Grid (caBIG) initiative, 2) Various CRI focused standards and harmonization bodies such as the Clinical Data Interchange Standards Consortium (CDISC), Health Level 7 (HL7), and the Biomedical Research Integrated Domain Group (BRIDG), 3) The creation and growth of clinical trial data registries, and 4) other NIH Road map initiatives such as the “re-engineering the clinical research enterprise” program which preceded the CTSA initiative and spawned several research projects that focused on management needs of the clinical research domain.

Additionally, the past five to ten years has seen the emergence of a growing body of literature describing central challenges to the national clinical research enterprise and the corresponding benefits that could result from addressing those challenges through, in part, effective integration of biomedical informatics and clinical/translational research workflows.
Methods

Given the early stage of knowledge in this domain, we employed a qualitative methodologic approach to identify the range of challenges and opportunities that face the CRI domain as well as defining its scope. Multiple phases were employed to approach the pertinent issues from various perspectives and achieve triangulation of findings. The particular methods and output associated with each phase are illustrated at a high-level in Fig 1, and described in greater detail in the sections that follow. The University of Cincinnati’s Institutional Review Board (IRB) approved this study.

Phase (1) Identify Community-perceived Challenges and Opportunities in CRI

Data Collection

We conducted a semi-structured, expert moderator facilitated group discussion during the 2006 AMIA annual symposium. As described in greater detail in the manuscript presenting a preliminary analysis of this phase’s findings, the session took place in a hotel meeting room on Nov 14, 2006, during the final 30 minutes of the Clinical Research Informatics Working Group Business meeting, following a series of presentations on various ongoing national and international CRI initiatives by expert panelists who then participated in the discussion along with other attendees.

To encourage broad participation in advance of the session, participant recruitment involved posting invitations about the planned session on multiple AMIA mailing lists as well as posting flyers announcing the session around the meeting area of the annual session hotel. Prior to starting the session, all prospective participants were notified of the planned study and provided with an IRB-approved information sheet describing the study methods and how any data being collected would be used for research purposes. Consent to have their comments recorded and analyzed in a confidential manner was implied if they remained and participated, though participation was not required for those who remained to listen. Further, to avoid any sense of pressure to participate, all were informed that they would have another opportunity to participate after the session via the working group’s e-mail list. The facilitated discussion session was conducted by a researcher with experience conducting such group interviews (PE) using a semi-structured interview technique, and the discussion was audio-taped. This approach was used to allow for the emergence of unanticipated relevant issues while still ensuring coverage of topics of anticipated importance based upon prior discussions within our working group.

The discussion session began with a brief reminder of the CRI focus and a mention of the purpose of the discussion. This was followed by open-ended questioning of the group about their views on three major issues: (1) the challenges facing CRI; (2) the opportunities facing CRI; and (3) the role that AMIA could play in this domain. Open-ended follow-up questions were asked as needed to catalyze deeper discussion of issues that the investigators considered of high importance in satisfying the aims of this study. Only when these techniques yielded little or no response were closed-ended questions used to further explore issues of potential importance as anticipated by the investigators before the session. Microphones were provided to all participants to clearly capture each utterance. The audio-taped recording was professionally transcribed verbatim and the audio file was converted to an MP3 to facilitate audio-review during the analysis phase by the investigators as needed for verification of the transcript. Field notes were manually recorded during the session by an experienced investigator (PP).

Data Analysis

Field notes and verbatim transcripts of the approximately 25-minute recording were analyzed independently by two
reviewers (PE, PP). A sentence or phrase in the transcripts served as a unit of analysis for coding purposes. Content analysis was performed using a grounded theory approach to identify emergent themes. Descriptions for each unit of analysis were based upon participants’ utterances whenever possible. After each researcher conducted his independent review of the transcripts and notes, the investigators entered an iterative process beginning with discussion of their descriptions of the data. This led to agreement on broad categorizations that could be directly linked to the raw data. As a final step, the categorizations were organized into common themes.

**Phase (2) Follow-up E-mail Discussion with CRI Professionals**

**Data Collection**
Following phase 1, we proceeded to collect additional feedback from the broader CRI community by initiating a discussion on the AMIA CRI Working Group e-mail list. The members of this group had an opportunity to respond to an e-mail message that requested their feedback to questions phrased just as they were during the in-person discussion described in Phase 1 above. Beginning on Feb 1, 2007, the members of this group received three separate e-mail requests sent by the authors. These requests also included an IRB-approved notification that responses posted to the e-mail list would be analyzed for research purposes. Direct responses to the requests or responses to the postings of other respondents were collected for a period of 2 months at which point data collection was concluded.

**Data Analysis**
As with the data collected during Phase 1, the data collected during this phase were collated, and the investigators (PE, PP) first independently and subsequently collaboratively performed content analysis using the same approach described above for the data collected during Phase 1.

**Phase (3) Validation of Preceding Challenges and Opportunities by Domain Experts and Development of a Formal Definition of CRI**
To verify and validate the challenges and opportunities identified in the prior study phase, we used a combination of qualitative approaches referred to as member-checking and peer-debriefing, as described below:

**Data Collection**
For member-checking, a selected sample of participants from the session conducted during Phase 1 who are widely recognized as opinion leaders in the biomedical informatics and/or clinical research domain and who were willing to provide us with additional feedback (n = 8) were asked to comment upon whether the findings identified in the previous phases and presented below were representative of the key CRI issues being targeted by this study. For peer-debriefing, we identified additional opinion leaders in the fields of biomedical informatics and/or clinical research (n = 12) to comment on those findings. In total, we engaged 20 domain experts in this phase.

As in Phase 1, the subjects in this phase represented the academic, government, and industry sectors. First, we sent each subject the summary of our findings from Phase 2 and asked them to review the findings and commenting upon them while answering the following questions: (1) “Do the current categories/themes capture the range of major challenges and opportunities facing the CRI domain today? If not, what do you feel is missing?”; and (2) “Are there any other additions/changes you would recommend?” Responses from these subjects were returned via e-mail to the study investigators.

In a separate, second e-mailing, we sent this same group of domain experts a preliminary version of a definition of CRI and an accompanying list of exemplary CRI focus areas. These materials were initially developed by members of the AMIA CRI Working Group, including the authors, and they were subsequently modified based upon our thematic findings and insights collected during the preceding phases before delivery to our subjects. The domain experts were asked to critique and comment upon the definition and list of CRI focus areas in an open-ended fashion and reply with their comments to the investigators.

**Data Analysis**
Responses to each of the two intervention steps conducted during this phase were aggregated separately for qualitative analysis conducted using the same techniques employed in Phases 1 and 2. For the data collected during the first findings-validation step of this phase, the investigators focused particularly on identifying any new or different findings that might justify modification of the results of prior phases. For the data collected during the second definition-critique step of this phase, the investigators each focused on identifying the range of comments made and modifying the definition and focus-area list as appropriate.

**Phase (4) Aggregation and Summary Analyses of Findings from the Preceding Phases**
For the last phase of our study, we aggregated the primary findings from the preceding phases for a final, summary review and analysis stage. Specifically, we took note of issues that were identified by more than one methodology and particularly attempted to identify any new overarching themes or new organizational representations through which the findings’ meaning could be augmented. After independent review of the findings and development of any new thematic coding and reorganization, we discussed our findings, resolving any discrepancies, and consolidated our individual findings from this phase into a common set of emergent items, summary interpretations, and aggregated presentation that arose from the cumulative findings.

**Results**
In the following subsections, we summarize the findings of our multiphase study, including the issues, challenges and opportunities identified, and present them organized by the phases in which they were generated.

**Phase (1) Identify Community-perceived Challenges and Opportunities in CRI**
Forty-six people were in attendance at the facilitated discussion and given the opportunity to contribute to the discussion, and 22 actively participated. To maintain the confidentiality of this relatively small group of participants, detailed demographics cannot be presented. However, we can report that participants represented a broad range of backgrounds and perspectives including: academic informaticians, clinical and
translational investigators, governmental funding agency representatives, leaders of major health and research informatics initiatives and professional organizations, pharmaceutical and health IT industry professionals, and health IT investors. All participants were directly involved in biomedical informatics-related and/or clinical research-related activities as a major focus of their jobs. Participants responded spontaneously to the open-ended questions posed, with little need for the interviewer to prompt responses via the use of follow-up questions. As such, there was minimal need for closed-ended follow-up to discussion points. In total, 3 pages of field notes and 7 single-spaced pages of verbatim transcript of the approximately 25-minute recording were analyzed by two reviewers (PE, PP) using the methods described earlier.

Analysis of the data yielded numerous distinct categories of comments related to challenges, concerns, and opportunities facing CRI. From these categories, twelve themes emerged that together encompass all comments made during the discussion. While some data elements could be assigned to unique categories, many represented more than one concept and were therefore cross-categorized. For example, if a comment addressed the need for education about data standards, it was coded as relating to both education and standards. The 12 broad themes included: research planning and conduct; data access and integration; educational needs; fiscal and administrative issues; policy issues; leadership and coordination; recruitment issues; scope of CRI; socio-technical issues; standards; workflow; lessons not learned.

Phase (2) Follow-up E-Mail Discussion with CRI Professionals
Forty-three e-mail responses were received during this study phase in response to investigator initiated messages that requested comments from AMIA CRI working group members concerning their perceptions of prevailing challenges and opportunities facing the clinical research informatics domain. Among these, 4 (10%) messages were direct responses to investigator prompts, while an additional 39 (90%) messages were related to the initial prompts, but were not direct responses to them. Thematic analyses of these messages demonstrated that they spanned a complete spectrum of topic areas, including all major themes described in the prior study phases. Notably, no new categories or themes arose during this phase that were not already captured during the Phase 1 study. The most common themes under which findings from these messages were categorized were 1) the need for leadership and coordination of CRI research and development efforts; 2) the need for increased CRI education for informaticians and researchers; 3) the need for more widespread adoption of data encoding or interchange standards; and 4) the pervasive nature of socio-organizational factors that prevent the ready adoption of CRI focused IT platforms.

Phase (3) Validation of Preceding Challenges and Opportunities by Domain Experts, and Development of a Formal Definition of CRI
Among the 20 participants in this phase of the study, 10 agreed that the initial list captured the range of issues, while another 10 participants commented how the current list of themes and categories could be expanded, clarified, or reorganized. Analysis of these comments led us to augment and modify our preliminary findings identified during Phase 1 of our study and presented previously. While many of the items expressed were broadly covered in existing categories, some felt that certain items were of sufficient importance to require their own category, and there were a few items that were newly added.

Examples of items previously identified but deserving of expansion included: regulatory over-interpretation as an impediment to research conduct; challenges regarding not just data access and integration, but also analysis of data across sites and across types of data (i.e., biological and clinical); the need to standardize workflow and reporting requirements across sites and sponsors; and the need to have CRI professionals involved in setting the CRI agenda at leadership levels within institutions and in regulatory/governmental agencies. In addition to these and other important augmentations to our preliminary findings, new elements emerged during this validation process. These largely centered around the new theme we entitled “CRI innovation and investigation”, a theme that included such subcategories as 1) “research in CRI is critical to advancing the field”; 2) “research in CRI often is secondary to fulfilling current needs”; 3) “there is a need for more funding to advance CRI theory, not just practice”; and 4) the related “need to recognize CRI professionals’ efforts in promotion and tenure considerations.”

Therefore, this validation phase led to modification of some of our previously identified themes and categories as well as adding one new theme and related subcategories to the set of findings identified. The resultant 13 identified and validated themes along with their constituent categories and example quotations are summarized in Table 1. Narrative descriptions of these final 13 emergent themes derived from these three phases are summarized here:

- **Research planning and conduct**: it became clear from most of our participants’ responses throughout all phases of our study that explicit and effective connections between CRI platforms and capabilities, and the planning or execution of clinical research efforts, were lacking if not nonexistent. This lack of coordination between informaticians and research investigators or staff was seen as a primary impediment to the realization of the potential benefits in terms of research productivity or capacity afforded by CRI methods and tools.
- **Data access, integration, and analysis**: Study participants felt that a multitude of organizational, policy-based, and practical factors collectively made the ability to access comprehensive and/or integrative data sets throughout the clinical Research Spectrum, difficult to achieve. This issue was particularly evident in discussions surrounding the secondary use of clinical data in support of research activities.
- **Educational needs**: Study participants identified a major need to educate informaticians, clinical research investigators/staff, and senior leadership concerning the theory and practice of CRI. Such education was thought to be necessary to ensure appropriate expectation management; adoption/use of CRI related methods or tools; and the allocation of appropriate resources to accomplish organizational aims.
# Summary Table of Emergent Themes, Their Underlying Categories, and Representative Quotations

<table>
<thead>
<tr>
<th>Themes</th>
<th>Underlying Categories</th>
<th>Representative Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>Research planning and conduct</td>
<td>Improved research planning tools&lt;br&gt;Advanced clinical research design&lt;br&gt;Tools cumbersome/poorly integrated&lt;br&gt;Need for advanced analysis tools/methods</td>
<td>“...the issue is that the tools are cumbersome and poorly integrated with the workflow of patient care, and so you can’t recruit either the investigators or the patients.”&lt;br&gt;“(Investigators) over-promise because the work is too hard to accomplish with the tools they are given.”&lt;br&gt;“Our computing power is great, yet we are basically using the calculating capability that was available to Fisher and his contemporary 80 yrs ago, without real appeal to the “knowledge level.””</td>
</tr>
<tr>
<td>Data access, integration, and analysis</td>
<td>Poor data access&lt;br&gt;Lack of data integration&lt;br&gt;Secondary use of data issues&lt;br&gt;Incorporate research into clinical systems&lt;br&gt;Need improved analysis across sites</td>
<td>“I think one of the largest challenges institutionally is data access and integration, and that’s the barrier for the investigator.”&lt;br&gt;“...the interface between the electronic medical record and the clinical research is critical.”&lt;br&gt;“In the era of integration between clinical systems and clinical trials systems, there remain some interesting unsolved problems.”</td>
</tr>
<tr>
<td>Educational needs</td>
<td>Educate students, investigators, clinicians about CRI&lt;br&gt;Educate those working in CRI&lt;br&gt;Educate informaticians about CRI&lt;br&gt;Need for cross-discipline education&lt;br&gt;Educate senior leadership about CRI</td>
<td>“...it might be worth considering what is the training required, because it is a tremendous leap for a DBA to understand the regulations for running a clinical trial. It is a tremendous leap for an electronic health records guy. CIOs and academic centers do not get it. No offense. This is a big leap here.”&lt;br&gt;“(We should) educate some of our informatics colleagues...about the obstacles and roles and needs that we have in this research environment.”</td>
</tr>
<tr>
<td>Fiscal and administrative issues</td>
<td>Research billing challenges&lt;br&gt;Costs of research software&lt;br&gt;Improved business processes needed&lt;br&gt;Lack of incentives for adopting research tools/need to demonstrate ROI for CRI solutions</td>
<td>“We are struggling with the appropriateness of tracking research costs and research charges and making sure that we are compliant with that and what is research and what is not. So, there is a whole other set of business processes on the investigator’s side that need to be somehow looped and coordinated with processes on the study side”&lt;br“So one of the big things driving this is the complete lack of money or incentive to move beyond paper.”</td>
</tr>
<tr>
<td>Regulatory and policy issues</td>
<td>International CRI activities&lt;br&gt;Regulatory frameworks&lt;br&gt;Political obstacles&lt;br&gt;Regulation mis/over-interpretation as impediments to progress&lt;br&gt;Security and privacy issues</td>
<td>“A broad issue is the international nature of clinical research, especially clinical trials; for example, clinical trials in developing countries where the informatics infrastructure and the regulatory and ethical oversight are sometimes not as well developed—trials having to respond to a patchwork of national, regional, and international regulations and responsible agencies”&lt;br&gt;“...overly conservative or incorrect interpretations of a regulation can become an inadvertent impediment to clinical research”</td>
</tr>
<tr>
<td>Leadership and coordination</td>
<td>Need for coordinated CRI agenda&lt;br&gt;Need for setting practical goals&lt;br&gt;Need for coordination between initiatives/among stakeholders&lt;br&gt;Desire for leadership/guidance from professional organization&lt;br&gt;Need for coordination between regulators and CRI community&lt;br&gt;Need to have CRI professionals represented in institutional leadership&lt;br&gt;No clear channel for CRI decision making within academic institutions</td>
<td>“...(the part) AMIA can play a role in is keeping up with all of this...so creating this portal that allows people from this group to be able to contribute and to go and understand what is going on in this space.”&lt;br&gt;“(We need to encourage) open comment from all the different perspectives so we are hearing from the investigator, the vendor, the institution, the NCI and other NIH institutes; it would be fabulous to get that input (about CRI initiatives).”&lt;br&gt;“...absent leadership from CRI sitting at the table in the Regent’s meeting and at the Hospital board meeting and in the corridors of the Capitol, there is little hope of the software developer curing the problem.”&lt;br&gt;“...no group across academic medical institutions fully ‘own’ the problem of solving clinical research informatics.”</td>
</tr>
<tr>
<td>Recruitment issues</td>
<td>Ineffective subject recruitment&lt;br&gt;Current tools make recruitment difficult&lt;br&gt;Lost opportunities to recruit&lt;br&gt;Investigator recruitment challenges&lt;br&gt;Not maximizing existing clinical information systems for recruitment</td>
<td>“Recruitment is the single biggest challenge. The investigators always over-promise. They do feasibility assessment. They try to target as best they can but there is under-performance that causes them to have to do rescues mid-way through. They have to close down sites, start up new sites. They cannot get a good handle on how to predict or estimate, or enhance and augment recruitment.”&lt;br&gt;“So the issue again rests on what can we do to make it easier so that clinicians are interested in becoming investigators.”</td>
</tr>
<tr>
<td>Scope of CRI</td>
<td>Recognize CRI is about more than informatics for clinical trials&lt;br&gt;Include support for nursing research&lt;br&gt;Include research partners in agenda (e.g., pharma, government)</td>
<td>“(We should) expand our vision from clinical trials to encompass all of clinical research, because I think this is going to be critical as we move forward. It is not just the trials. It is the outcomes, the biomarkers. It is the epidemiology studies.”&lt;br&gt;“Nurses use a wide range of research techniques, and I think we need to be investigating those as well.”</td>
</tr>
</tbody>
</table>


- **Fiscal and administrative issues**: Study participants voiced three primary concerns relative to the ability of the CRI community to overcome fiscal and administrative barriers to the implementation and use of informatics platforms in the clinical research setting, namely: 1) the inability to overcome billing compliance issues (e.g., disambiguation of standard-of-care versus research related billable activities in an automated manner)—which are perceived as a primary motivator for the adoption of CRI platforms in many large organizations; 2) a lack of consistent funding mechanisms for persistent CRI infrastructure (e.g., beyond episodic grant funding mechanisms); and 3) the absence of demonstrable return on investment (ROI) or alternative fiscal models that may serve to justify the support of CRI platforms at the organizational level.

- **Regulatory and policy issues**: respondents voiced the opinion that the current regulatory environment is confusing, misunderstood, and often contradictory. They specifically stated that there is an essential conflict between the regulatory frameworks faced by the clinical and translational sciences and public or governmental demand for safer or more effective therapies. This conflict exists due to the desire to systematically protect patient privacy and confidentiality that often constrains the ability to apply informatics techniques to develop and evaluate novel therapies or otherwise conduct clinical research.

- **Leadership and coordination**: the study participants were concerned about the current state of coordination of current national, regional, and local-scale CRI research programs which appear to be focusing on the same essential aims. Many participants felt that such programs included policy and fiscal barriers that prevented coordination—thus exacerbating this particular problem.

- **Recruitment issues**: participants raised various concerns around challenges to human subject recruitment for clinical research studies as well as challenges related to recruitment of site investigators for multicenter studies.

### Table 1 • continued.

<table>
<thead>
<tr>
<th>Themes</th>
<th>Underlying Categories</th>
<th>Representative Quotations</th>
</tr>
</thead>
<tbody>
<tr>
<td>9. Socio-organizational issues</td>
<td>Research/clinical missions not integrated/coordinated</td>
<td>“There are some obvious technical challenges (to providing data integration and access), but...one of the biggest challenges coming is actually integrating the research mission into the academic health care environment and ensuring that (removal of) these barriers that we have to data and systems and use of them for research is able to happen.”</td>
</tr>
<tr>
<td>10. Standards</td>
<td>Need for CRI data standards, models</td>
<td>“…properly representing medical concepts and the right terminology in order to access anything down stream which includes real world data from insurance and electronic medical records, databases, as well as other knowledge bases.”</td>
</tr>
<tr>
<td></td>
<td>Apply clinical standards to research</td>
<td>“…sponsors/pharma don’t have standards and make each center create things according to each sponsor separately, like invoices, budgets, data entry, SAE reporting, and subject tracking.”</td>
</tr>
<tr>
<td></td>
<td>Need ways to span biological to clinical ontologies</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Need to standardize nontechnical institutional and sponsor requirements</td>
<td></td>
</tr>
<tr>
<td>11. Workflow</td>
<td>Integrate tools into workflow</td>
<td>“...we all complain about the time it takes for IRB review.”</td>
</tr>
<tr>
<td></td>
<td>Inefficiency of research processes, need for effective models</td>
<td>“(Current systems suffer from) a complete lack of integration between patient care processes and clinical research.”</td>
</tr>
<tr>
<td></td>
<td>IRB (Institutional Review Board)/regulatory challenges</td>
<td>“While a lot of talk has been made about standards, (so) we can move data back and forth, what is actually happening is there’s no work processes that make that happen.”</td>
</tr>
<tr>
<td></td>
<td>Need to consider users needs</td>
<td>“The paucity of such thinking and research (about transformational approaches and theory) in CRI means that information and technology management advances have been subservient to the current process. There has been little of the disruption and leapfrogging that would change the environment.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“the leadership in CRI gets bogged down in “pet projects” because that is the nature of their funding, and disruptive theory and models (that help advance CRI), instead of being the raison d’être for their departments, actually is rather scary for them.”</td>
</tr>
<tr>
<td></td>
<td></td>
<td>“Academic promotion committees do not require or reward data sharing or data standardization efforts”</td>
</tr>
<tr>
<td>12. CRI innovation and investigation</td>
<td>Research in CRI is critical to advancing field</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Research in CRI often secondary to fulfilling current/practical needs</td>
<td></td>
</tr>
<tr>
<td></td>
<td>CRI needs to focus on transformative advances rather than just supporting current practice</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Academic promotion process does not reward CRI practice</td>
<td></td>
</tr>
<tr>
<td>13. Lessons not learned</td>
<td>Concern about repeating clinicalinformatics mistakes</td>
<td>“…the problems and the hopeful solutions (being proposed) are so analogous to the issues that have swirled around the electronic health record here at AMIA for the last 15 or 20 years or so, and I would be a little more comfortable if more was said about what went wrong with that. Because to start all over again with the same good hopes and the same stuff about standards and we will make software and all the rest of it, it didn’t cut it the last time, so I am hoping that the lessons from the past will be incorporated into the work on the clinical research (agenda).”</td>
</tr>
<tr>
<td></td>
<td>Need for best practices</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Under-use of current tools/best practices</td>
<td></td>
</tr>
</tbody>
</table>
In particular, participants noted that while well validated approaches and associated informatics platforms designed to improve subject recruitment to clinical trials have been developed, the ability to readily adopt and deploy such tools and approaches in “real world” setting is impeded by both a lack of applicable standards and a set of policy-based barriers surrounding the use of clinical data to enable such applications.

- **Scope of CRI**: most of our study participants felt that the nature of CRI and its definition was still very ambiguous and often misunderstood. As an outgrowth of this finding, the definition of CRI presented in our results was developed and validated. Of interest in this context was a concern raised by several respondents that the areas of clinical research and clinical trials were often conflated when discussing such definitions and scope, while in fact clinical trials are a subset of the broader clinical research paradigm.

- **Socio-organizational issues**: the topic of socio-organizational issues, as raised by the study participants, spanned a spectrum from local policy-based objections; to the need for greater integration between institutional clinical, research, and educational mission areas, to the lack of sufficient resources as needed to implement robust CRI research, development, and support efforts. The most critical theme identified surrounding the topic of sociotechnical issues was the fact that technical barriers are not the sole issue that CRI researchers and practitioners must deal with to be successful in their field, and that overcoming incompatible and often times misinformed sociotechnical barriers and objections to informatics interventions in the clinical research domain was as important if not more important than technical hurdles.

- **Standards**: Study participants felt that the adoption of appropriate data coding and representation standards in the CRI domain was greatly impeded by both the large number of competing standards available to select from, as well as the absence of appropriate standards or approaches capable of supporting information exchange between the basic science, clinical research, and clinical practice domains.

- **Workflow**: Study respondents indicated that they felt many if not most CRI platforms suffered from a lack of integration with existing or optimized clinical research workflows, thus impeding their adoption. Respondents repeatedly called for an increased focus on the development and validation of workflow optimization and informatics intervention strategies for the clinical research domain. In particular, these types of workflow issues arose in the context of the intersection of clinical research activities and clinical care.

- **CRI innovation and investigation**: many of our study participants felt that a community-accepted understanding of the essential or high priority scientific problems and research foci needed to advance the discipline of CRI was lacking. There were clear and repeated calls for an emphasis in the CRI community on “research-on-research” to advance the discipline. Participants also felt that greater coordination across national research efforts and funding agencies was needed to catalyze and support the definition of key CRI research problems and corresponding research agendas.

- **Lessons not learned**: The last thematic finding pertains to multiple comments by study participants that there were similarities between the preceding 12 thematic areas and challenges or opportunities that had been faced during the evolution of the clinical, public health, and bio-informatics subdisciplines. The consensus opinion of the respondents was that we should ensure the CRI community takes heed of the lessons generated by such similar experiences in other biomedical informatics communities, rather than replicate their experiences.

### Defining Clinical Research Informatics

As mentioned above, a key finding identified throughout our study phases was that members of the biomedical informatics community felt that a formal definition of CRI was both necessary and lacking. Among the 20 participants in this phase of the study that we asked to review and comment upon our definition of CRI, 4 agreed that the definition was adequate, while 16 had suggestions for improving it. Most of these centered around the insertion of additional key words to capture the activities engaged in by CRI professionals as well as helpful semantic clarifications. While many also grappled with the fact that CRI clearly overlaps considerably with the closely associated field of translational research informatics (TRI), the general consensus was that this definition need not necessarily capture and represent the entire clinical and translational research informatics spectrum by combining CRI and TRI into a single definition. It was also concluded that this definition should build upon rather than attempt to redefine the fields of biomedical informatics and clinical research that have been previously defined by others. Finally, respondents agreed that an accompanying list of high-level representative CRI focus areas was helpful to augment this definition and had some suggestions for augmenting the preliminary list presented. The result is following definition for CRI:

#### Clinical Research Informatics (CRI)

**CRI** is the subdomain of biomedical informatics concerned with the development, application, and evaluation of theories, methods, and systems to optimize the design and conduct of clinical research and the analysis, interpretation, and dissemination of the information generated.

Augmenting this definition is the following non-exhaustive, but illustrative list of CRI focus areas and activities:

- evaluation and modeling of clinical and translational research workflow
- social and behavioral studies involving clinical research professionals and participants
- designing optimal human-computer interaction models for clinical research applications
- improving and evaluating information capture and data flow in clinical research
- optimizing research site selection, investigator, and subject recruitment
- knowledge engineering and standards development as applied to clinical research
- facilitating and improving research reporting to regulatory agencies
- enhancing clinical and research data mining, integration, and analysis
• integrating research findings into individual and population level health care
• knowledge integration across clinical and research information systems
• defining and promoting ethical standards in CRI practice
• educating researchers, informaticians, and organizational leaders about CRI
• driving public policy around clinical and translational research informatics

Phase (4) Aggregation and Summary Analyses of Findings from the Preceding Phases

Based upon our iterative assessment of the findings from the preceding phases, it became clear that the 13 emergent themes of CRI challenges and opportunities could be logically organized and aggregated into higher-level themes. It was also clear that these overarching themes crossed stakeholder groups and appeared to fall into place along two-dimensional axes as follows.

First, the emergent higher-level themes involved three major groupings relevant to the advancement, practice, and leadership aspects of CRI. These included:

1. **CRI academics and advancement**, which encompasses the themes of: educational needs; scope of CRI, CRI innovation and investigation.
2. **Practice of CRI**, which encompasses the themes of: research planning and conduct; data access, integration and analysis; recruitment issues; workflow; standards.
3. **Society and leadership aspects of CRI**, which encompasses the themes of: regulatory and policy issues; leadership and coordination; socio-organizational issues; financial and administrative issues.

4. **Lessons not learned** arose as an overarching theme that was relevant to and spanned all the above higher-level themes.

Second, it was evident that these thematic groupings were logically related to various stakeholder groups affected by these findings that roughly spanned geographic scales from:

1. **Individuals**, including Researchers and IT/Informatics Professionals;
2. **Organizational entities**, including academic and private sector institutions;
3. **National and International entities**, such as funders of research and informatics, regulatory bodies, and governmental agencies.

A graphical representations of the themes and findings reported above and organized by these higher-level thematic and stakeholder dimensions is presented in Fig 2.

**Discussion**

Those involved and familiar with efforts at the intersection of clinical research and biomedical informatics identify CRI as an important new subdiscipline that faces a range of critical challenges and opportunities. To our knowledge, these challenges and opportunities have not previously been formally assessed or organized to illustrate the degree to which they span the various stakeholder groups identified above. Particularly at this relatively early stage in the development of a new domain, the availability of a systematically and formally developed set of challenges, opportu-

---

**Figure 2.** Overview of identified themes organized into higher-level groupings by scope, and applied across the groups of stakeholders to which they apply. Of note is the fundamental and cross-cutting nature of the frequently articulated theme labeled as *Lessons not learned.*
nities, and definition should help to facilitate further action, prioritize resourcing, and spur planning and agenda development to advance the field.

**Defining CRI**

While presenting a formally vetted, community-derived definition of the CRI domain is one part of this study’s results that collectively describe the CRI domain, we feel this outcome is notable as it serves to delineate the domain’s scope and augment the list of major challenges and opportunities identified. Of course, as with any effort to define a complex domain, this definition will almost certainly not suit everyone’s needs or desires. Indeed, even among our varied respondents, there were conflicting sentiments about the right balance between generality and specificity for the definition. Another area of debate related to whether translational research informatics (TRI), a closely related and similarly emergent discipline practiced by many who focus on the full spectrum of clinical and translational research informatics, should be captured under the definition of CRI. While it was our conclusion that, just as clinical and translational research are distinct but closely related domains, so too are CRI and TRI, it is also clear that multiple intersections exist between CRI, TRI, and other informatics and research-related subdisciplines that may support the argument that they are somewhat inextricable from each other.

By examining the relationships between CRI and other areas of biomedical informatics domains of knowledge and practice as depicted in Fig 3, we can better understand how informatics subdomains concerned with aspects of basic and early translational science (e.g., Bioinformatics), clinical practice (e.g., clinical informatics), and population health (e.g., public health informatics) overlap with and complement CRI. For instance, an overlap exists between CRI and the related but distinct field of Translational Bioinformatics on the T1 end and with clinical and public health informatics that exists at the T2 end of the translational research spectrum. We feel that this illustration also serves to complement the definition of CRI presented above by addressing some of the “fuzzy” boundaries that exist between the complementary subdomains of biomedical informatics, and which were noted by subjects in all phases of our study.

**Implications of Findings**

Beyond definitional issues, an examination of the categorizations and themes that emerged from this study is also significant in that it enumerates major issues facing CRI, issues that have broad implications for the domain. When these findings are considered within the current state of biomedical informatics knowledge and practice, several additional points and suggestions for action arise as particularly noteworthy:

- **While CRI is a unique, emergent subdiscipline of biomedical informatics, it is also exceptionally cross-cutting in nature.** As alluded to in the discussion above, CRI, at least as much and perhaps more than other informatics subdomains, draws upon or contributes to other informatics domains by its very nature. Indeed, as a subdiscipline concerned with critical aspects of the broad translational science spectrum, CRI necessarily relates to and advances through coordination and cross-fertilization of theories and methods with related subdomains spanning the basic and clinical aspects of biomedical informatics.

- **CRI is not simply concerned with IT platforms for the clinical research setting and should not be thought of that way by those involved in clinical research or biomedical informatics enterprises.** Unfortunately, as voiced by many of our concerned participants, this perception is the norm in many settings. It is the perspective of the authors that biomedical informatics and IT can be situated in a translational spectrum. At one end of that spectrum, the scientific pursuit of biomedical informatics is concerned with the creation and evaluation of novel theories, frameworks, models, and practices as needed to address essential biomedical information management needs, while at the other end of the same spectrum, IT is concerned with the operationalization and support of such theories, frameworks, models, and practices in “real world” or “produc-
There is a critical need for the development of targeted policy and/or research agendas, driven by the CRI Community, to inform the structure and function of research programs and funding mechanisms at all levels. There are many stakeholders with interests in the advancement of CRI, but, as our findings make evident, there is concern that the policies and research initiatives that have helped to drive the CRI agenda have often been parts of other primary initiatives. While helpful, there is widespread concern that such efforts are often too narrowly focused (e.g., only on technical development and infrastructure to support existing workflows), are frequently under-resourced, and often result in uneven development of products and tools at the expense of more fundamental CRI knowledge, best-practices, and expertise. With a growing group of informaticians focused on the CRI space, there is a critical need for this CRI community to drive a strategic policy and research agenda to help avoid haphazard and ad hoc solutions that while well-intentioned may result in further fragmentation and worsening of existing challenges.

trusted and widely accessible mechanisms for the sharing of knowledge, best practices, and technologies targeting the theory and application of CRI must be established. The CRI community is concerned that all too often knowledge about CRI innovations, best practices, and solutions are maintained in inaccessible silos. While there are many reasons including a lack of incentives for sharing resources and knowledge, one potential solution is the creation and maintenance of trusted mechanisms for sharing such knowledge, insights and resources among the CRI Community. It is critical that such solutions not only be widely accessible, but that they extend beyond the technical products of CRI, given the importance of theory and methods to achieving success across CRI challenges.

Greater advocacy for the rationalization of regulatory frameworks with clinical research informatics as a key component for consideration alongside clinical and public health concerns must be established. Few issues elicited more concern from our respondents than those related to regulatory challenges. While there was a clear understanding of, appreciation for, and even championing of the need for strong regulations surrounding such issues as information privacy and data reporting, there was also a concern that such regulations have often been developed with little or no consideration for the implications they may have for the clinical research enterprise. The result of these regulations has at times been to impair the research enterprise’s ability to advance discovery, thereby leading to costly delays and failures that our healthcare sector cannot afford to sustain. Here again, there is an urgent need for involvement of the CRI community to inform the rationalization of regulations and policies, particularly related to personal and population level health data, so that research information needs and considerations are taken into account. Such efforts must be undertaken with the goal of assuring privacy and safety while not inadvertently impairing our ability to advance the very scientific advancements that will ultimately lead to improvements in health.

Professional organizations, such as AMIA, should play a role to help overcome certain CRI challenges and catalyze the development of broad CRI solutions by bringing together key stakeholders and serving as impartial, coordinating entities. While independent groups, organizations and agencies have and certainly can continue to achieve significant advances through independent or partially coordinated efforts geared at addressing some of the CRI challenges identified previously, these efforts have and will often continue to be limited by the nature of the groups sponsoring them. What is necessary to address many of the challenges identified in the context of advancing CRI as a professional discipline is a venue for discussion, coordination and collaboration surrounding existing and new initi-
atives involving the full range of stakeholders interested in the CRI agenda.

Limitations
There are several limitations to note related to the work presented in this manuscript in addition to those already mentioned above. While the researchers’ status as participant-observers of the CRI-community provide them with insights into the domain, we acknowledge that this can also inadvertently introduce some biases into our qualitative analyzes. In addition, the self-selected and limited convenience sample of participants, particularly in phases one and two of our studies, may not represent the composition of the broader CRI community. It is also possible that even with our multiphase, multimethod approach, we did not capture all the challenges and opportunities facing the CRI domain. Despite such limitations, we feel that our use of qualitative multieexpert validation techniques mitigate these shortcomings and enhance the validity of our findings.

Conclusions
Clinical Research Informatics (CRI) has emerged as a distinct subdiscipline of biomedical informatics and one that while still maturing is faced with many challenges and opportunities. We believe that the findings presented in this report provide empiric support for advancing the field of CRI beyond its current, nascent state toward a more mature discipline by providing the contextual underpinnings that can inform a cohesive and systematic cycle of research, development, and evaluation. By engaging in such a transition at this juncture, the CRI community has the opportunity to avoid the roadblocks and impediments experienced by other informatics subdisciplines during their own, similar, formative stages. It is our hope that this grounded scoping of the CRI space as well as cataloguing of the challenges and opportunities facing the CRI domain will serve to catalyze the pace and scale of CRI advancement, thereby enabling improvements in human health through advances in clinical and translational science.

References