Clinical Research Informatics: Survey of Recent Advances and Trends in a Maturing Field

P. J. Embi
The Ohio State University, Columbus, OH, USA

Summary

Objectives: To provide a survey of the field of clinical research informatics (CRI), focusing in particular on significant developments over the past 3 years and the insights they provide about the progress and state of the field.

Methods: An iterative "scoping review" of the literature published in scientific journals and conference proceedings that are relevant to CRI, from late 2009 to early 2013.

Results: 212 articles were identified, and 64 were selected to illustrate recent advances in the field. Based on those, six categories of CRI activity were identified: Data and Knowledge Management, Discovery and Standards; Clinical Data Re-Use for Research; Researcher Support and Resources; Participant Recruitment; Patients/Consumers and CRI; Policy, Regulatory and Fiscal Matters.

Conclusions: This survey demonstrates that the field of CRI has matured and is now well established. The ongoing work is essential to overcome many of the challenges the clinical research enterprise is facing and more work is needed. Even as work continues to establish necessary infrastructure, methods, and best practices, CRI researchers should strive for more rigorous study designs to evaluate the impacts of the work in the field. There is little doubt that the field is poised for rapid growth, and that the CRI literature will continue to reflect that growth in years to come.

Keywords

Clinical research; medical informatics; clinical trials; translational science; review literature as topic

Introduction

Clinical research is critical to the advancement of medical science and public health. Conducting such research is a complex resource and information-intensive endeavor involving multiple actors, workflows, processes, and information resources. The need to accelerate clinical research and to solve the information-related problems inherent to it have brought increased attention to a field at the intersection of clinical research and biomedical informatics throughout governmental, academic, and private sectors [1-8]. The result is the emergence and relatively rapid expansion of a biomedical informatics sub-discipline focused on clinical research, referred to as Clinical Research Informatics (CRI). As the field has matured, work has been done to define its scope and the range of challenges and opportunities it addresses [9].

This maturity of CRI has resulted in a steady increase in work across the CRI space, and such efforts have begun to result in significant improvements in the quality and efficiency of clinical research [5, 10]. Indeed, recent years have seen the emergence of national and international research and policy efforts to foster advances in CRI. CRI professionals have addressed the inherent challenges and opportunities that motivate the sub-discipline, including those in particular areas of research such as Cancer [2, 8, 11, 12] and more broadly those focused on reengineering the research enterprise to accelerate such science. Examples include initiatives such as the National Institutes of Health's (NIH) Clinical and Translational Science Award (CTSA) (3, 4), as well as grants from institutes and agencies such as the National Library of Medicine and the Agency for Healthcare Research & Quality (AHRQ) to drive advances in research data methods and resources for CRI-related works, including comparative effectiveness and health services research [13-15].

In addition to initiatives focused on advancing the science and practice of CRI, investments in Health information technology (IT) adoption spurred on by incentives and "meaningful use" of EHRs are also playing a role in CRI advances, if less directly. While initially primarily focused on improving patient care, interoperable EHRs and improvements in standards-based data sharing hold great promise for the streamlined collection and reuse of data and information from clinical care to improve public health and research – to create the learning health system [16].

Just as biomedical informatics approaches and resources are essential to realizing the potential of such systems to improve clinical care, so are the methods, theories and tools of the CRI domain critical to realizing the potential of such systems to enable discovery through acceleration and enhancement of clinical research. As a result, advances in CRI continue to accelerate.

One marker that CRI has emerged as a distinct and relevant domain is the growth of a significant literature focused on this topic. Indeed, publications describing advances in the field are increasing and serve as evidence that the field is maturing. With that in mind, the sections that follow are an attempt to represent *some* of the excellent work in the CRI field over the past few years. However, it should be noted that any attempts to survey this field would fall short of providing a comprehensive view. Instead my intent is to present the reader with exemplary cases and an overview of this important and rapidly advancing domain of biomedical informatics.

Methods

Searching for relevant manuscripts in a relatively new domain wherein subject headings and keywords are not well established requires an iterative approach, sometimes referred to as a *scoping review* [17]. In this case, a structured literature search was augmented with a snowballing method involving reviewing citations in identified articles as well as articles recommended by other experts in the domain to whom the author reached out for recommendations.

The literature review for this survey of "recent" activity in the CRI domain involved combinations of MESH terms and keywords as well as date restrictions, with the goal of identifying relevant articles that covered a range of issues actively being studied. For the literature review, we utilized the following query in PubMed: ("Biomedical Research"[Mesh] NOT "Genetic Research"[Mesh]) NOT "Translational Research"[Mesh]) AND "Informatics"[Mesh] AND "2009/01/01"[PDat] : "2013/02/01"[Pdat]. At that time, this search returned 298 articles, among which 101 were judged by the author to be CRI-relevant. In addition, another 111 articles were identified via a combination of recommendations by colleagues with expertise in CRI; by reviewing references in identified articles; and via ad hoc searches using keywords such as: "Clinical Trials"; "Clinical Research"; "Informatics"; "Translational"; "Data Warehouse"; "Recruitment".

Again, this is not a systematic review, so the articles selected merely represent the author's attempt to highlight relevant and/or representative recent works in CRI, and should not be considered an exhaustive list of the field's relevant literature. Also worth noting is that this review only includes the field of "clinical research informatics" but not the closely related, overlapping, but distinct sub-domain of translational bioinformatics.

Upon identifying the articles via the above-described approach, the author categorized them into descriptive areas of activity. Under each category, some of the relevant articles are highlighted. The article concludes with some notable CRI-related events and a brief discussion of future directions.

Results

Based on the methodology described above, 212 articles were identified, and 64 were selected to illustrate recent advances in the field. Based on these articles, several categories of CRI activity can be identified:

- Data and Knowledge Management, Discovery and Standards
- Clinical Data Re-Use for Research
- Researcher Support and Resources
- Participant Recruitment
- Patients/Consumers and CRI
- Policy, Regulatory and Fiscal Matters

Data and Knowledge Management, Discovery, and Standards

Data and knowledge management, including data warehousing, has become central to many activities essential to the clinical research process such as hypothesis generation, protocol development and feasibility analyses, and data extraction for analytical and reporting purposes. Biomedical research has benefited from the foundational work in data warehousing from other sectors, and has built upon that work with developments and methods focused on enabling data storage, retrieval and integration for clinical and translational science. One example is the NIH-supported "informatics for integrating biology and the bedside" platform, also known as i2b2 [18]. Created for cohort finding, easy querying for a range of research uses, and easy data-mart creation, the i2b2 platform has been adopted by over 60 centers worldwide [19], is extensible to enable integration of tools for analysis and other research activities [20], and offers the capability to conduct queries and identify cohorts across instances for inter-institutional research [21, 22]. Infrastructural developments like this are essential resources for enabling the clinical and translational research advances that concern CRI.

Similar to clinical data warehousing efforts, activities are also underway to enable the storage, discovery, and sharing of biospecimens and tissues relevant to the conduct of clinical and translational sciences. For instance, as part of the National Cancer Institute's caBIG program, Crowley and colleagues described the development of a grid-based Cancer Tissue Information Extraction System (caTIES) for coding surgical pathology reports and retrieving tissue specimens [23]. Implemented at several cancer centers across the US, caTIES leverages existing NLP methods, caGrid computing and security frameworks, and query visualization methods to enable data sharing, graphically-driven concept-based text mining, and a regulatory and security model for multi-center collaboration.

Management and discovery of data and knowledge derived from clinical research activities themselves also represent an important area of CRI activity. One area of activity involves efforts to ensure clinical trials are systematically registered in accessible databases like clinicaltrials.gov [14]. Leveraging such a system, Tasneem and colleagues demonstrated the utility of a resource they designed to enable aggregated analyses, organized by specialties, of content registered in clinicaltrials.gov [24]. Of course, the ability to retrieve, review and analyze the actual results of clinical trials holds great promise to improve their veracity, transparency, and value to human health. To that end, Sim and colleagues reported on approaches to enable the "human studies database project" by developing the ontology of clinical research (OCRe) that facilitates federated, machine-readable, study design and results information for interventional and observational studies [25]. Such CRI resources Can enable large-scale data mining, synthesis and re-analysis of trial data, thereby greatly enhancing the value of individual studies for biomedical advances.

Fundamental to this category and to the following ones are the myriad initiatives related to the development, application and adherence to standards. Standards efforts in CRI include the development of technical standards, data standards and ontologies, and other clinical research standards, many of which are essential to the management of data and knowledge in and across repositories [26]. Standards efforts related to the capture, management and exchange of clinical research data include those led by groups such as the Clinical Data Interchange Standards Consortium (CDISC), HL7, the FDA [27]. Among such efforts are the Biomedical Research Integrated Domain Model (BRIDG), the Integrating Healthcare Enterprise (IHE)

initiatives, and others concerned with different aspects of facilitating research processes using healthcare systems [28, 29]. Other efforts relate to the development of content standards for the transmission of trial information to the FDA (i.e. the Study Data Tabulation Model) [27], that enable data exchange between clinical and research information systems (i.e. the Operational Data Model) [30], and representations for eligibility criteria [31]). Still other examples involve the development of standards for sharing data about non-clinical resources, such as the development by Tenenbaum and colleagues of the "Biomedical Research Ontology" (BRO) to enable the semantic search and discovery of research resources across sites [32]. There are many more such efforts ongoing, and several others have written recently about the status of advances, remaining challenges, and the opportunities they present in CRI [33-37].

Clinical Data Re-Use for Research

As mentioned above, leveraging the ever-increasing collection of digitized clinical data in EHRs for the purpose of systematic learning, evidence generation, and healthcare improvement is a goal that our healthcare system must reach. Informatics activities are central to such data reuse, but there are many challenges both technical and socio-organizational that must be overcome in order to realize the potential benefits, and there are many ways to re-use clinical data for research.

As noted in the previous section, there are ongoing efforts to develop standards and resources for re-using clinical data in research studies. For example, the Strategic Health IT Advance Research Projects (SHARP) Program includes initiatives (SHARPn) specifically focused on extending the use of EHR data for "secondary" purposes such as health outcomes improvement, biomedical research, and epidemiologic monitoring of the nation's health, all via the development of open-source services and platforms [38]. Another set of efforts involves data re-use for drug safety monitoring and surveillance, such as the FDA's Sentinel system [39] and the Adverse Drug Events Spontaneous Triggered Event Reporting (ASTER) study [23]. Still others have leveraged the aforementioned IHE initiatives to enable

single-source data entry in the clinical setting through the integration of "Retrieve Form for Data Capture" (RFD) with the IHE "Clinical Research Document" (CRD) to repurpose EHR data for research-specific electronic care report forms (eCRF) [40] and more general clinical research uses [28].

A key data reuse activity for research involves identifying patient cohorts from EHR-derived data sets, for recruitment (which is addressed in greater detail below) as well as for other research processes, but this can be challenging when key data are buried in narrative text. In their 2012 article, Carroll and colleagues demonstrated that their phenotype algorithm to identify patients with rheumatoid arthritis from EHRs was effective when ported to other institutions with distinct EHR systems [41]. By comparing the results of their algorithm with the gold-standard of physicians' manual review of charts to identify the patients, they were able to show that NLP approaches used against EHR-derived data performed nearly as well, with AUCs between 92% and 95% vs. 97%, respectively. Retraining the logistic regression model improved results, and all were shown to be better than billing code count thresholds, demonstrating that electronic phenotyping algorithms can allow for rapid identification of certain populations across sites with different EHRs, NLP systems, and little retraining. Indeed, leveraging EHR data to identify specific patient cohorts and phenotypic representations has also proven to be effective in other recent studies at the intersection of EHR and genomic data – findings that have implications for CRI as well as translational bioinformatics [42-44].

Despite those promising findings, recent studies in the US, Europe and Africa have also shown that while certainly useful for identifying certain cohorts and even monitoring outcomes in certain conditions [45], data primarily collected for clinical purposes may not always be optimal for research uses unless attention is given during its collection with such an end in mind [45-47]. For instance, when assessing whether quality measurements for performance monitoring can be automatically derived from EHRs, Parsons and colleagues found some limitations [48]. Based on their chart reviews of over 4000 patients across 57 practices, they found that

automated EHR-derived quality measures underestimated actual performance, documentation varied between sites, and results often did not reflect the accurate number of patients who actually got preventive measures. Other recent studies have similarly demonstrated that data collected for a general clinical care purpose may not be of adequate quality for research when compared to that systematically collected in a dedicated registry for a given patient population [46]. Even when evaluating accuracy for non-research reporting needs, such as automated quality reporting for "meaningful use", significant variability is noted [49], emphasizing challenges and the need for further CRI developments to ensure effective data re-use.

Not surprisingly, efforts to address issues about streamlined collection, accuracy and completeness of clinical data for reuse purposes represent another active area of CRI research [50]. In one notable study, Wright and colleagues reported on their use of a clinical alert system using inference rules to notify providers of undocumented problems in an EHR problem list, and showed that such an approach can indeed improve problem documentation [51].

Researcher Support & Resources

Many ongoing efforts in CRI revolve around the development of resources for use by researchers as they conduct their day-to-day research activities. A recent survey of academic health centers in the US reveals that there has been progress regarding key research resource availability at the institutional level, as compared to earlier surveys dating back five to seven years. Significant progress in four main areas revealed increases in: (i) the use of IT in research compliance, such as conflicts of interest, research budgeting, and reporting to the Institutional Review Board (IRB); (ii) the use of IT for electronic data capture (EDC) requirements related to clinical studies and trials of different size; (iii) the use of data repositories for the repurposing of clinical care data for research; and, (iv) the IT infrastructure needs and support for research collaboration and communication [52]. The authors attributed this to the availability of more robust commercial and "open-source"

solutions, coupled with new research initiatives (e.g., CTSA) and regulatory requirements necessitating such infrastructure.

Among specific resources that have taken hold are those enabling coordination of research support requests at an institutional level. The StarBRITE system at Vanderbilt University is one such system [39]. It is a homegrown content management system that provides a single, integrated Web-based portal for investigator to request support while also serving the administrative needs of the institution. Such systems thereby have the potential to improve how institutions "care for" and "monitor" investigators and the research enterprise. By leveraging such researcher portals, approaches have also been developed to help researchers navigate the regulatory and organizational processes that often slow down scientific progress [53]. CRI efforts have also been reported on the knowledge engineering approaches specific to informing the design and enhancing the usability of such researcher support portals (54).

Another basic research activity supported by recent CRI developments is related to research data collection. The Research Electronic Data Capture (REDCap) system is a prime example of an easy-to-use, metadata-driven methodology and workflow process that provides an easy to manage and use electronic data capture (EDC) solution for common research needs [29]. Now in use at over 200 sites internationally, and including over 9200 studies and over 18,000 end-users, REDCap has become a standard tool for many of the EDC needs of researchers across the globe. Where implemented it has effectively diminished the one-off database development needs that used to consume resources, present security risks, and slow scientific progress. Building upon the REDCap foundation, collaborators have even extended the system to areas where EDC has traditionally been challenged. By combining REDCap tools and a custom data synchronization application, Borlawsky and colleagues enabled distributed data collection in rural Appalachian sites with limited Internet connectivity [55].

While stand-alone EDC systems are and will continue to be essential, data collection from EHRs is also increasing, particularly as the lines between clinical care and research become blurred. As such, another area of CRI

activity relates to the development of research registries in EHR systems. Indeed, even as some research demonstrates the superiority of registries over standard clinical data collection for certain uses [46], other studies demonstrate that registry data can even be equal to that collected during phase III RCTs, potentially obviating the need for some clinical trials, if properly designed and used [56].

Another resource that is becoming common for research management is the Clinical Trial Management System, sometimes referred to as the Clinical Research Management System (CRMS). While there are many such systems in the vendor-space, some have recently been the subject of CRI research. One called TRITON, integrates tissue management components from the caTISSUE suite with CRMS functionalities to provide an integrative platform where researchers can manage their research from tissues to research participant management [57]. As CRMS use expands, next steps will include their integration or at least connectivity with EHRs and specimen repositories for more robust and cohesive research management.

As research support systems take hold and investments continue to grow, so will the demand to demonstrate their utility and cost-effectiveness. Such assessments will require new ways to track and evaluate the evolution of research activities and collaborations across multiple scales, and some CRI investigators are beginning to establish baseline assessments through the use of network analysis approaches across an institution [58]. These and other approaches can be expected to increase as CRI advances continue.

Participant Recruitment

The identification and recruitment of participants for clinical research is well recognized as a major impediment to the efficient conduct of research studies. The vast majority of clinical trials suffer from significant delays in achieving enrollment targets, and this in turn delays medical advances and is costly [59]. As such, informatics approaches to overcome the issues that plague recruitment have long been an area of CRI activity [60, 61]. Recent work has built upon prior work in a variety of ways.

Over the period of this review, there have been several intervention studies to overcome

recruitment delays. Dugas and colleagues studied a form of EHR-driven recruitment alerts in a German hospital, and demonstrated that daily email alerts sent to clinicians with the list of potentially eligible subjects yielded 329 enrollments for seven distinct trials over a 10-month period [38]. In another study, Herasevich and colleagues addressed the issue of patient recruitment to studies of time-sensitive conditions (i.e. septic shock). By automatically paging a trial coordinator when EHR features indicated the patient was developing shock, they effectively doubled enrollment [62]. Weng and colleagues evaluated the utility of using a registry versus a data warehouse to recruit participants for a diabetes clinical trial [63]. They found that even while each option had its own benefits and challenges, the use of a data warehouse yielded more accuracy (31% vs. 6.6%) and higher participant recruitment (30 vs. 14) in a shorter time period (59 vs. 74 working days) than did the registry.

While alert-based approaches appear to show great promise to enhance recruitment in a variety of settings, the issue of alert fatigue often arises as a concern. In a study published last year, alert fatigue was evaluated as part of a randomized controlled trial of the clinical trial alert approach [64]. Alert fatigue did appear to occur as response rates to the clinical trial alert declined gradually over time, but they remained at a relatively high level, 36% even after 36 weeks of exposure to alerts, suggesting it wasn't as severe as some had worried.

In addition to the ongoing work to leverage EHRs for recruitment, other systems designed to accelerate participant identification, matching, and enrollment have also been the subject of CRI activity. ResearchMatch, a national registry for volunteers interested in learning about research opportunities is one such solution. With over 30,000 volunteers, in use by 76 institutions, and with about a 20% acceptance rate by registrants being contacted for a study, this registry is a major resource in the CRI domain [65]. Other resources dedicated to improving recruitment rates include some data-driven cohort matching systems within institutions [66, 67].

As many studies designed to enhance recruitment have shown, one of the challenges to leveraging information systems for alerting is the problem of precisely and accurately matching patient data to clinical

trial eligibility criteria. One reason is that eligibility criteria are often not written to be computer-interpretable, a matter covered in detail in a review of the topic in 2010 [68]. Systems are now being developed and tested to extract eligibility criteria and represent them in computer-interpretable formats to improve matching [69], even as other approaches to improve computability of criteria during the protocol authoring process are ongoing.

Patients/Consumers and CRI

The role of patients and/or consumers of healthcare in research has traditionally been limited to a relatively passive one whereby the "subject" is often a willing participant in a process that is largely controlled by the research professionals. This has recently changed. First, more attention is being paid to the informed consent process, a key part of traditional patient engagement that is sometimes not optimized for all participants. Recent studies have applied more scrutiny to the often poor readability of informed consent documents and to the use of technologies to improve informativeness [70, 71].

Further engagement of patients is taking place through the use of increasingly ubiquitous mobile technologies to collect data in ways that are more appealing and appear to be at least as accurate as traditional methods [72]. In one recent study, patients were engaged to contribute information regarding their medications and side-effects via a personal health record, thereby correcting the underlying clinical database from which research activities such as hypothesis generation or even participant eligibility might be determined [73].

Even more direct involvement in the research process by patients and/or consumers of healthcare is taking place. Disease-specific social networks are leading to community-initiated studies, and the fact that the patient community drives such studies is also leading to more rapid accruals and study completion [70, 71].

Policy, Regulatory and Fiscal Matters

The research enterprise and individual research studies exist in the often complex world alongside clinical care. As such, the realities of clinical care, privacy concerns, and limited resources are often positioned opposite to the needs of researchers, and research enterprise more broadly. In addition, regulatory issues related specifically to research, such as adherence to FDA guidance on the use of computer systems for research data management and federal regulations like the 21 CFR Part 11 rule, must be addressed and therefore drive many ongoing CRI efforts.

Essential to the functioning of the research enterprise is abiding by the ethical and regulatory tenets of society. A major one in the case of CRI relates to safeguarding data privacy even while enabling discovery and advancing science. Recent work by Malin and colleagues exemplifies work to provide the CRI community approaches to address some of the pressing privacy issues related to data sharing [74]. However, as the Institute of Medicine recently (IOM) pointed out, there have long been major flaws with existing privacy regulations, including inconsistencies between the US HIPAA privacy rule and the so-called 'common rule' governing aspects of research, variable interpretations across organizations, and inadequately staffed IRBs with limited ability to effectively interpret the rules. These issues lead in many cases to overly conservative decisions that impede research progress without enhancing privacy, and the IOM has called to rework the privacy rule so that, "Rather than seeing privacy and research as conflicting values, policy makers can improve both"[30]. The recent release in 2013 of revisions to the HIPAA privacy rule appears to address some of these issues.

Another important topic relates to fiscal matters that impact upon clinical research and hence the informatics endeavors to support them. On the one hand, there are remarkable investments being made in the US to accelerate health IT adoption and meaningful use with an eye toward creating a "learning health system" [16]. In addition, there are major investments ongoing to advance patient-centered outcomes research, comparative effectiveness research, and pragmatic clinical trials, all of which expect an information infrastructure will exist to enable such activities [75-77]. On the other hand, the clinicians and healthcare providers who are expected to perform high-quality data collection for such research are not compensated for performing such activities. As a result, some in the CRI community have called for approaches to encourage clinician participation in research activities. One proposal argued for a "relative research unit" counterpart to the "relative value unit" upon which clinical care is measured and compensated, to similarly measure and reward the additional work required to advance research alongside clinical care [78].

Overall, there are myriad policies, regulations, and fiscal matters that currently impede our collective ability to advance work at the intersection of biomedical informatics and clinical research, and these must be addressed as some have begun to call for [79]. Work must also continue to advance foundational research in the area of CRI to enable the innovations needed in clinical and translational science [80-82]. In addition, work must begin to train a new generation of clinicians and healthcare leaders in biomedical informatics, with CRI as a central component, even as more CRI-focused informaticians are also needed (83).

Conclusion

As this survey demonstrates, the field of CRI is broad and rapidly advancing. Indeed, there are many other examples and categories of CRI activity that were not possible to adequately cover in this brief survey. There are also other signs that the field is maturing, including the recent publication of the first textbook dedicated to CRI [84], the establishment and growth of an academic meeting dedicated to the field (85), and journal special issues dedicated to work in CRI [86, 87].

Advances across a range of data, knowledge, and information efforts focused on clinical research are becoming common across the research enterprise landscape. Even still, rapid acceleration of science via informatics approaches is not yet commonplace, and CRI interventions still appear too often as anecdotes or quasi-experimental studies, with methodologically rigorous studies (e.g. randomized, controlled studies) of CRI interventions remaining rare. As the field matures, more research on CRI efforts will be needed to demonstrate the successes and value of CRI interventions.

Given the rapid advances in biomedical discoveries, the growth of the human population, and the escalating costs of health care, the need to accelerate and improve clinical research is essential. As such, the field of CRI has emerged as critical to solving many of the current challenges faced by clinical researchers and the research enterprise. There is little doubt the field is poised for continued and rapid growth, and that growth will be reflected in the biomedical informatics literature in the years to come.

References

1. The Cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community. Stud Health Technol Inform 2007;129(Pt 1):330-4.

2. Kakazu KK, Cheung LW, Lynne W. The Cancer Biomedical Informatics Grid (caBIG): pioneering an expansive network of information and tools for collaborative cancer research. Hawaii Med J 2004 Sep;63(9):273-5.

3. Zerhouni EA. Translational and clinical science--time for a new vision. N Engl J Me. 2005 Oct 13;353(15):1621-3.

4. Zerhouni EA. Clinical research at a crossroads: the NIH roadmap. J Investig Med 2006 May 1;54(4):171-3.

5. Payne PR, Johnson SB, Starren JB, Tilson HH, Dowdy D. Breaking the translational barriers: the value of integrating biomedical informatics and translational research. J Investig Med 2005 May;53(4):192-200.

6. Sung NS, Crowley WF, Jr., Genel M, Salber P, Sandy L, Sherwood LM, et al. Central challenges facing the national clinical research enterprise. JAMA 2003 Mar 12;289(10):1278-87.

7. Oster S, Langella S, Hastings S, Ervin D, Madduri R, Phillips J, et al. caGrid 1.0: an enterprise Grid infrastructure for biomedical research. J Am Med Inform Assoc 2008 Mar-Apr;15(2):138-49.

8. Saltz J, Oster S, Hastings S, Langella S, Kurc T, Sanchez W, et al. caGrid: design and implementation of the core architecture of the cancer biomedical informatics grid. Bioinformatics 2006 Aug 1;22(15):1910-6.

9. Embi PJ, Payne PR. Clinical research informatics: challenges, opportunities and definition for an emerging domain. J Am Med Inform Assoc 2009 May-Jun;16(3):316-27.

10. Chung TK, Kukafka R, Johnson SB. Reengineering clinical research with informatics. J Investig Med 2006 Sep;54(6):327-33.

11. Oster S, Langella S, Hastings S, Ervin D, Madduri R, Phillips J, et al. caGrid 1.0: An Enterprise Grid Infrastructure for Biomedical Research. J Am Med Inform Assoc 2008 Jan 1;15(2):138-49.

12. The Cancer Biomedical Informatics Grid (caBIG): infrastructure and applications for a worldwide research community. Medinfo 2007;12(Pt 1):330-4.

13. Collen MF. Clinical research databases--a historical review. J Med Syst 1990 Dec;14(6):323-44.

14. Gillen JE, Tse T, Ide NC, McCray AT. Design, implementation and management of a web-based data entry system for ClinicalTrials.gov. Stud Health Technol Inform 2004;107(Pt 2):1466-70.

15. Holve E, Segal C, Lopez MH, Rein A, Johnson BH. The Electronic Data Methods (EDM) forum for comparative effectiveness research (CER). Med Care 2012 Jul;50 Suppl:S7-10.

16. Friedman CP, Wong AK, Blumenthal D. Achieving a nationwide learning health system. Sci Transl Med [Research Support, U.S. Gov't, Non-P.H.S.]. 2010 Nov 10;2(57):57cm29.

17. Levac D, Colquhoun H, O'Brien KK. Scoping studies: advancing the methodology. Implement Sci 2010;5:69.

18. Murphy SN, Weber G, Mendis M, Gainer V, Chueh HC, Churchill S, et al. Serving the enterprise and beyond with informatics for integrating biology and the bedside (i2b2). J Am Med Inform Assoc 2010 Jan 1;17(2):124-30.

19. Kohane IS, Churchill SE, Murphy SN. A translational engine at the national scale: informatics for integrating biology and the bedside. J Am Med Inform Assoc [Research Support, N.I.H., Extramural]. 2012;19(2):181-5.

20. Segagni D, Ferrazzi F, Larizza C, Tibollo V, Napolitano C, Priori SG, et al. R Engine Cell: integrating R into the i2b2 software infrastructure. J Am Med Inform Assoc 2011 Jan 24.

21. Weber GM, Murphy SN, McMurry AJ, Macfadden D, Nigrin DJ, Churchill S, et al. The Shared Health Research Information Network (SHRINE): a prototype federated query tool for clinical data repositories. J Am Med Inform Assoc 2009 Jan 1;16(5):624-30.

22. Anderson N, Abend A, Mandel A, Geraghty E, Gabriel D, Wynden R, et al. Implementation of a deidentified federated data network for population-based cohort discovery. J Am Med Inform Assoc 2011 Aug 26.

23. Brajovic S, Piazza-Hepp T, Swartz L, Dal Pan G. Quality assessment of spontaneous triggered adverse event reports received by the Food and Drug Administration. Pharmacoepidemiol Drug Saf 2012 Jun;21(6):565-70; discussion 71-2.

24. Tasneem A, Aberle L, Ananth H, Chakraborty S, Chiswell K, McCourt BJ, et al. The Database for Aggregate Analysis of ClinicalTrials.gov (AACT) and Subsequent Regrouping by Clinical Specialty. PLoS One 2012 Apr 16;7(3):e33677.

25. Sim I, Carini S, Tu S, Wynden R, Pollock BH, Mollah SA, et al. The human studies database project: federating human studies design data using the ontology of clinical research. AMIA Summits Transl Sci Proc 2010 Jan 1;2010:51-5.

26. Richesson RL, Krischer J. Data standards in clinical research: gaps, overlaps, challenges and future directions. J Am Med Inform Assoc 2007;14(6):687-96.

27. Kush RD, Helton E, Rockhold FW, Hardison CD. Electronic health records, medical research, and the Tower of Babel. N Engl J Med 2008;358(16):1738-40.

28. Holzer K, Gall W. Utilizing IHE-based electronic health record systems for secondary use. Methods Inf Med 2011;50(4):319.

29. Fridsma DB, Evans J, Hastak S, Mead CN. The BRIDG project: a technical report. J Am Med Inform Assoc 2008 Mar-Apr;15(2):130-7.

30. Breil B, Kenneweg J, Fritz F, Bruland P, Doods D, Trinczek B, et al. Multilingual medical data models in ODM format–a novel form-based approach to semantic interoperability between routine health-care and clinical research. Appl Clin Inf 2012;3:276-89.

31. Weng C, Tu SW, Sim I, Richesson R. Formal representations of eligibility criteria: A literature review. J Biomed Inform 2010;43(3):451.

32. Tenenbaum JD, Whetzel PL, Anderson K, Borromeo CD, Dinov ID, Gabriel D, et al. The Biomedical Resource Ontology (BRO) to enable resource discovery in clinical and translational research. J Biomed Inform 2011 Mar;44(1):137-45.

33. Kong YM, Dahlke C, Xiang Q, Qian Y, Karp D, Scheuermann RH. Toward an ontology-based framework for clinical research databases. Journal of biomedical informatics. 2011 Mar;44(1):48-58.

34. Szalma S, Koka V, Khasanova T, Perakslis ED. Effective knowledge management in translational medicine. J Transl Med 2010 Jan 1;8:68.

35. Richesson RL, Nadkarni P. Data standards for clinical research data collection forms: current status and challenges. J Am Med Inform Assoc 2011 May 12;18(3):341-6.

36. Shaw M, Detwiler LT, Noy N, Brinkley J, Suciu D. vSPARQL: a view definition language for the semantic web. J Biomed Inform 2011 Mar;44(1):102-17.

37. Wynden R, Weiner M, Sim I, Gabriel D, Casale M, Carini S, et al. Ontology mapping and data discovery for the translational investigator. AMIA Summits Transl Sci Proc 2010;2010:66-70.

38. Rea S, Pathak J, Savova G, Oniki TA, Westberg L, Beebe CE, et al. Building a robust, scalable and standards-driven infrastructure for secondary use of EHR data: The SHARPn project. J Biomed Inform 2012;45(4):763-71.

39. Behrman RE, Benner JS, Brown JS, McClellan M, Woodcock J, Platt R. Developing the Sentinel System—a national resource for evidence development. N Engl J Med 2011;364(6):498-9.

40. El Fadly A, Rance B, Lucas N, Mead C, Chatellier G, Lastic P-Y, et al. Integrating clinical research with the healthcare enterprise: from the RE-USE project to the EHR4CR platform. J Biomed Inform 2011;44:S94-S102.

41. Carroll RJ, Thompson WK, Eyler AE, Mandelin AM, Cai T, Zink RM, et al. Portability of an algorithm to identify rheumatoid arthritis in electronic health records. J Am Med Inform Assoc 2012 Mar 28.

42. Denny JC, Ritchie MD, Basford MA, Pulley JM, Bastarache L, Brown-Gentry K, et al. PheWAS: demonstrating the feasibility of a phenome-wide scan to discover gene-disease associations. Bioinformatics 2010 May 1;26(9):1205-10.

43. Ritchie MD, Denny JC, Crawford DC, Ramirez AH, Weiner JB, Pulley JM, et al. Robust replication of genotype-phenotype associations across multiple diseases in an electronic medical record. Am J Hum Genet 2010 Apr 9;86(4):560-72.

44. Kho AN, Pacheco JA, Peissig PL, Rasmussen L, Newton KM, Weston N, et al. Electronic medical records for genetic research: results of the eMERGE consortium. Sci Transl Med 2011 Apr

20;3(79):79re1.

45. Dregan A, Toschke MA, Wolfe CD, Rudd A, Ashworth M, Gulliford MC. Utility of electronic patient records in primary care for stroke secondary prevention trials. BMC Public Health 2011;11:86.

46. Sagreiya H, Altman RB. The utility of general purpose versus specialty clinical databases for research: warfarin dose estimation from extracted clinical variables. J Biomed Inform 2010 Oct;43(5):747-51.

47. Kiragga AN, Castelnuovo B, Schaefer P, Muwonge T, Easterbrook PJ. Quality of data collection in a large HIV observational clinic database in sub-Saharan Africa: implications for clinical research and audit of care. J Int AIDS Soc 2011;14:3.

48. Parsons A, McCullough C, Wang J, Shih S. Validity of electronic health record-derived quality measurement for performance monitoring. J Am Med Inform Assoc 2012 Mar 09.

49. Kern LM, Malhotra S, Barron Y, Quaresimo J, Dhopeshwarkar R, Pichardo M, et al. Accuracy of Electronically Reported "Meaningful Use" Clinical Quality Measures: A Cross-sectional Study. Ann Intern Med 2013 Jan 15;158(2):77-83.

50. Herzberg S, Rahbar K, Stegger L, Schafers M, Dugas M. Concept and implementation of a computer-based reminder system to increase completeness in clinical documentation. Int J Med Inform 2011 Jun;80(5):351-8.

51. Wright A, Pang J, Feblowitz JC, Maloney FL, Wilcox AR, McLoughlin KS, et al. Improving completeness of electronic problem lists through clinical decision support: a randomized, controlled trial. J Am Med Inform Assoc 2012 Feb 03.

52. Murphy SN, Dubey A, Embi PJ, Harris PA, Richter BG, Turisco F, et al. Current state of information technologies for the clinical research enterprise across academic medical centers. Clin Transl Sci 2012 Jun;5(3):281-4.

53. Pulley J, Harris P, Yarbrough T, Swafford J, Edwards T, Bernard G. An informatics-based tool to assist researchers in initiating research at an academic medical center: Vanderbilt Customized Action Plan. Acad Med 2010 Jan 1;85(1):164-8.

54. Payne PRO, Borlawsky TB, Rice R, Embi PJ. Evaluating the impact of conceptual knowledge engineering on the design and usability of a clinical and translational science collaboration portal. AMIA Summits Transl Sci Proc 2010 Jan 1;2010:41-5.

55. Borlawsky TB, Lele O, Jensen D, Hood NE, Wewers ME. Enabling distributed electronic research data collection for a rural Appalachian tobacco cessation study. J Am Med Inform Assoc 2011 Dec;18 Suppl 1:i140-3.

56. Shi Q, You YN, Nelson H, Allen MS, Winchester D, Stewart A, et al. Cancer registries: a novel alternative to long-term clinical trial follow-up based on results of a comparative study. Clin Trials 2010 Dec 1;7(6):686-95.

57. Payne PRO, Borlawsky TB, Stephens W, Barrett MC, Nguyen-Pham T, Greaves AW. The TRITON Project: Design and Implementation of an Integrative Translational Research Information Management Platform. AMIA Annu Symp Proc 2010 Jan 1;2010:617-21.

58. Nagarajan R, Lowery CL, Hogan WR. Temporal evolution of biomedical research grant collaborations across multiple scales--a CTSA baseline study. AMIA Annual Symposium proceedings / AMIA Symposium AMIA Symposium 2011;2011:987-93.

59. Kitterman DR, Cheng SK, Dilts DM, Orwoll ES. The prevalence and economic impact of low-enrolling clinical studies at an academic medical center. Acad Med 2011 Nov;86(11):1360-6.

60. Embi PJ, Jain A, Clark J, Bizjack S, Hornung R, Harris CM. Effect of a clinical trial alert system on physician participation in trial recruitment. Arch Intern Med 2005 Oct 24;165(19):2272-7.

61. Embi PJ, Jain A, Clark J, Harris CM. Development of an electronic health record-based Clinical Trial Alert system to enhance recruitment at the point of care. AMIA Annu Symp Proc 2005:231-5.

62. Herasevich V, Pieper MS, Pulido J, Gajic O. Enrollment into a time sensitive clinical study in the critical care setting: results from computerized septic shock sniffer implementation.J Am Med Inform Assoc 2011 Aug 16;18(5):639-44.

63. Weng C, Bigger JT, Busacca L, Wilcox A, Getaneh A. Comparing the effectiveness of a clinical registry and a clinical data warehouse for supporting clinical trial recruitment: a case study. AMIA Annu Symp Proc 2010 Jan 1;2010:867-71.

64. Embi PJ, Leonard AC. Evaluating alert fatigue over time to EHR-based clinical trial alerts: findings from a randomized controlled study. J Am Med Inform Assoc 2012 Jun 1;19(e1):e145-e8.

65. Harris PA, Scott KW, Lebo L, Hassan N, Lightner C, Pulley J. ResearchMatch: A National Registry to Recruit Volunteers for Clinical Research. Acad Med 2012 Feb;87(1):66-73.

66. Ferranti JM, Gilbert W, McCall J, Shang H, Barros T, Horvath MM. The design and implementation of an open-source, data-driven cohort recruitment system: the Duke Integrated Subject Cohort and Enrollment Research Network (DISCERN). J Am Med Inform Assoc 2011 Sep 23.

67. Pressler TR, Yen PY, Ding J, Liu J, Embi PJ, Payne PR. Computational challenges and human factors influencing the design and use of clinical research participant eligibility pre-screening tools. BMC Med Inform Decis Mak 2012;12:47.

68. Weng C, Tu SW, Sim I, Richesson R. Formal representation of eligibility criteria: a literature review. J Biomed Inform 2010 Jun 1;43(3):451-67.

69. Weng C, Wu X, Luo Z, Boland MR, Theodoratos D, Johnson SB. EliXR: an approach to eligibility criteria extraction and representation. J Am Med Inform Assoc 2011 Dec 16;18(Suppl 1):i116-i24.

70. Hall A, Rode M, Pagnoux C, Yacyshyn E. Patient-Driven On-Line Survey on Granulomatosis with Polyangiitis. Arthritis Rheum 2011 Nov 04;63(Suppl 10):2371.

71. Tweet MS, Gulati R, Aase LA, Hayes SN. Spontaneous Coronary Artery Dissection: A Disease-Specific, Social Networking Community-Initiated Study. Mayo Clinic Proceedings 2011 Sep;86(9):845-50.

72. Bellamy N, Wilson C, Hendrikz J, Whitehouse SL, Patel B, Dennison S, et al. Osteoarthritis Index delivered by mobile phone (m-WOMAC) is valid, reliable, and responsive. J Clin Epidemiol 2011 Mar;64(2):182-90.

73. Pulley J, Hassan N, Bernard G, Jirjis J, Schildcrout J, Robertson D, et al. Identifying unpredicted drug benefit through query of patient experiential knowledge: a proof of concept web-based system. Clin Transl Sci 2010 Jun 1;3(3):98-103.

74. Malin B, Benitez K, Masys D. Never too old for anonymity: a statistical standard for demographic data sharing via the HIPAA Privacy Rule. J Am Med Inform Assoc 2011;18(1):3-10.

75. Clancy CM, Slutsky JR. Commentary: a progress report on AHRQ's Effective Health Care Program. Health Serv Res 2007 Oct 1;42(5):xi-xix.

76. Collins FS. Reengineering translational science: the time is right. Sci Transl Med 2011 Jul 06;3(90):90cm17.

77. Lauer MS, Collins FS. Using science to improve the nation's health system: NIH's commitment to comparative effectiveness research. JAMA 2010 Jun 2;303(21):2182-3.

78. Embi PJ, Tsevat J. Commentary: the relative research unit: providing incentives for clinician participation in research activities. Acad Med [Editorial]. 2012 Feb;87(1):11-4.

79. Hersh W, Jacko JA, Greenes R, Tan J, Janies D, Embi PJ, et al. Health-care hit or miss? Nature 2011 Feb 17;470(7334):327-9.

80. DeMets DL, Califf RM. A historical perspective on clinical trials innovation and leadership: where have the academics gone? JAMA [Historical Article]. 2011 Mar 16;305(7):713-4.

81. Cantor MN. Translational informatics: an industry perspective. J Am Med Inform Assoc 2012 Mar;19(2):153-5.

82. Payne PRO, Embi PJ, Niland J. Foundational biomedical informatics research in the clinical and translational science era: a call to action. J Am Med Inform Assoc 2010 Nov 1;17(6):615-6.

83. Stead W, Searle J, Fessler H, Smith J, Shortliffe E. Biomedical Informatics: Changing What Physicians Need to Know and How They Learn. Acad Med 2010 Aug 12.

84. Richesson RL, Andrews JE. Clinical research informatics. London ; New York: Springer; 2012. Available from: http://dx.doi.org/10.1007/978-1-84882-448-5; http://proxy.ohiolink.edu:9099/login?url=http://dx.doi.org/10.1007/978-1-84882-448-5.

85. Sarkar IN, Payne PR. The joint summits on translational science: crossing the translational chasm. J Biomed Inform 2011 Dec;44 Suppl 1:S1-2.

86. Payne PRO, Embi PJ, Kahn MG. Selected papers from the 2011 Summit on Clinical Research Informatics. J Biomed Inform [Editorial]. 2011 Dec;44 Suppl 1:S54-5.

87. Calonge N. EDM forum supplement overview. Med Care 2012 Jul;50 Suppl:S1-2.

Correspondence to:
Peter J. Embi, MD, MS, FACP, FACMI
The Ohio State University
3190 Graves Hall
333 W. 10th Ave
Columbus, OH 43210, USA
Tel: +1 614 292 4778
Fax: +1 614 688 6600
E-mail: peter.embi@osumc.edu