Information Technology for Clinical, Translational and Comparative Effectiveness Research

Findings from the Yearbook 2015 Section on Clinical Research Informatics

C. Daniel[1,2], R. Choquet[1,3], Section Editors for the IMIA Yearbook Section on Clinical Research Informatics

[1] INSERM UMRS 1142, Paris, France
[2] Direction of Information Systems, AP-HP, Paris, France
[3] BNDMR, Necker Hospital for Children, AP-HP, Paris, France

Summary

Objectives: To select and summarize key constributions to current research and to select best papers published in 2014 in the field of Clinical Research Informatics (CRI).

Method: A bibliographic search using a combination of MeSH and free terms search over PubMed on Clinical Research Informatics (CRI) was performed followed by a double-blind literature review.

Results: The review process yielded four papers, illustrating various aspects of current research efforts done in the area of CRI. The first paper exemplifies the process of developping a domain ontology for integrating structured, unstructured, and signal data into a coherent structure for patient care as well as clinical research. In the second paper, the authors analysed in five sites' hospital information system environments in Germany the possibility of implementing a patient recruitment process and provided recommendations for the development of dedicated patient recruitment modules. The third paper describes the IMI EHR4CR project which developed an instance of a platform, providing communication, security and semantic interoperability services to the eleven participating hospitals and ten pharmaceutical companies located in seven European countries. The last paper describes the relation between health status severity and the availability of data in EHR systems. They demonstrate that it introduces a biasis in patient selection for clinical research.

Conclusions: Distributed research networks are growing in importance for clinical research and population health surveillance and current research demonstartes that different projects and initiatives could be well placed to deliver international scale solutions to enable the reuse of hospital EHR data to support clinical research studies. Selected articles demonstrate the potential of formal representation of multimodal and multi-level data in supporting data interoperability across clinical research and care domains. With the development of pragmatic research, designed with input from health systems and producing evidence that can be readily used to improve care, a key issue for "learning health care organizations" is to systematically assess the quality of their data.

Keywords

Medical informatics, biomedical research, clinical research, patient selection, phenotyping

Yearb Med Inform 2015;10:178-82
http://dx.doi.org/10.15265/IY-2015-030
Published online August 13, 2015

Introduction

Clinical research informatics (CRI) is a rapidly evolving sub-discipline within biomedical informatics that focuses on developing new informatics theories, tools, and solutions to accelerate the full translational continuum: basic research to clinical trials (T1), clinical trials to academic health center practice (T2), diffusion and implementation to community practice (T3), and "real world" outcomes (T4) [8]. It includes management of information related to clinical trials and involves informatics related to secondary use of routinely collected clinical data for research[1].

The goal of this section is to provide an overview of research trends and of "best" papers published in the past year that demonstrate excellent CRI relevant research.

Methodologies and tools supporting clinical research continue to be developed and evaluated in the various categories of CRI activity: data and knowledge management, clinical data re-use for research, methods in CRI, policy and perspectives, security, confidentiality and regulatory issues.

The selected papers this year illustrate current trends in the field that can be identified in the 2014 publications. The clinical research community has to address the important challenges related to the development of personalized medicine and patient-centered care coordination - this year's theme for the Yearbook. Although treating a single individual may be practical using current methods, providing personalized care routinely for all patients requires new informatics approaches for data and knowledge management. Another trend is that distributed research networks are growing in importance for clinical research and population health surveillance and some of them address the challenge of reusing electronic health record systems (EHRs) for accelerating clinical trials at a large international scale. In this context, a last trend is the current efforts towards consensus "best practices" for quality assessment of data collected during routine patient care in healthcare settings.

About the Paper Selection

A comprehensive review of published articles in 2014 addressing a wide range of issues for clinical research informatics was conductucted. The selection was performed by querying Pubmed/Medline (from NCBI, National Center for Biotechnology Information) with a set of predefined keywords: Biomedical Research, Clinical research, Medical research, Medical research, Pharmacovigilance, Patient Selection, Phenotyping, Genotype-phenotype associations, Data Collection, Epidemiologic Research Design, Epidemiologic Study Characteristics as Topic, Epidemiological Monitoring, Evaluation Studies as Topic, Clinical Trials as Topic, Feasibility Studies.

References addressing topics of other sections of the Yearbook, such as Translational Bioinformatics were excluded based on predefined exclusion keywords such as Genetic Research, Gene Ontology, Human Genome Project, Stem Cell Research or Molecular Epidemiology.

[1] http://www.amia.org/applications-informatics/clinical-research-informatics [Accessed: 24/05/2014].

Performed at the beginning of January 2015, the search yielded a total of 644 references. From this original set, a first subset of 269 references was considered according to relevancy to the CRI field and blindly reviewed by the two section editors. Based on title and abstract, the articles were classified into several CRI categories - data and knowledge management, clinical data re-use for research, methods in CRI, policy and perspectives and regulatory issues. Their contribution to the CRI category was rated as low, medium or high. Then, the two lists of references were merged, yielding 163 references that were classified as high contribution to CRI by at least one reviewer or medium contribution by both reviewers. The 163 references were reviewed by the two section editors jointly to select a consensual list of 18 candidate best papers. Following the IMIA Yearbook process, these 18 papers were peer-reviewed by editors and external reviewers (at least four reviewers per paper). Four papers were finally selected as best papers (Table 1). A content summary of these selected papers can be found in the appendix of this synopsis.

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2015 in the section 'Clinical Research Informatics'. The articles are listed in alphabetical order of the first author's surname.

Section
Clinical Research Informatics

- De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, Karakoyun T, Ohmann C, Lastic PY, Ammour N, Kush R, Dupont D, Cuggia M, Daniel C, Thienpont G, Coorevits P. Using electronic health records for clinical research: the case of the EHR4CR project. J Biomed Inform 2015 Feb;53:162-73.
- Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. BMC Med Inform Decis Mak 2014 Jun 11;14:51.
- Sahoo SS, Lhatoo SD, Gupta DK, Cui L, Zhao M, Jayapandian C, Bozorgi A, Zhang GQ. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. J Am Med Inform Assoc 2014 Jan-Feb;21(1):82-9.
- Schreiweis B, Trinczek B, Köpcke F, Leusch T, Majeed RW, Wenk J, Bergh B, Ohmann C, Röhrig R, Dugas M, Prokosch HU. Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. Int J Med Inform 2014 Nov;83(11):860-8.

Conclusion and Outlook

Active research is conducted worlwide to develop and deploy classical CRI techniques such as electronic data capture (EDC) and clinical data management system (CDMS) for clinical research and to promote the use of standards processes and formats [18].

In parallel, the widespread availability of electronic clinical data is enhancing the potential for supporting classical explanatory randomized controlled trials (RCTs) as well as observational studies and pragmatic clinical trials (PCTs) typically used to conduct comparative effectiveness research on multiple clinical sites and with broader eligibility criteria [3].

In the context of "Precision medicine" implying treating patients as individuals with specific characteristics at every level of the biological spectrum – a unique genome, proteome, metabolome, microbiome, etc and also a unique history of exposures, social history and personal preferences - novel approaches are proposed to augment institutional clinical research data warehouse.

The first selected paper, by Sahoo et al. [12] describes research in data and knowledge management focusing on integrating structured, unstructured, and signal data into a coherent structure for patient care as well as clinical research. This paper exemplifies the process of developping a formal domain ontology in the domain of epilepsy and seizure and its use to streamline the entry of patient information, to process clinical free text in patient records and to run federated queries to create patient cohorts across multiple study centers. Murphy et al. [10] describe an approach for accessing medical imaging examinations collected during routine clinical care and making them available to translational investigators. In the context of semantic interoperability between systems for both clinical research and secondary use, Jiang et al. [7] reported their preliminary efforts on harmonization of the SHARPn Healthcare Clinical Element Models (CEMs) that have been adopted by the Office of the National Coordinator (ONC) and data element extracted from the CDISC (Clinical Data Interchange Standards Consortium) templates.

The second selected paper, by De Moor et al. [9], presents the IMI EHR4CR platform, providing communication, security and semantic interoperability services to the eleven participating hospitals and ten pharmaceutical companies located in seven European countries. This paper demonstrates a scalable approach to interoperability between EHR systems and clinical research systems. In this context, Doods et al. [4] demonstrated that currently, not all information that is frequently used in site feasibility is documented in routine patient care and proposed a set of 75 data elements that, on the one hand are frequently used in clinical studies, and on the other hand are available in European EHR systems.

The third selected paper, by Schreiweis et al. [13] also describes an initiative in clinical data re-use for research in Europe. The authors analysed in five sites' hospital information system environments in Germany the possibility of implementing a patient recruitment process and provided recommendations for the development of dedicated patient recruitment modules.

Another study, by Tate et al. [15], present the use of UK primary care databases, which contain diagnostic, demographic and prescribing information for millions of patients geographically representative of the UK for identifying patients with a specified disease or condition and to investigate patterns of diagnosis and symptoms. Trifiro et al. [16] reviewed several international initiatives (e.g. EU-ADR, Sentinel, OMOP, PROTECT and VAESCO) based on the combined use of multiple healthcare databases for the conduct of active surveillance studies in the area of drug and vaccine safety and provide a summary of the potential, disadvantages, methodological issues and possible solutions of each of the initiative.

Another review, by Shivade et al. [14], presents a variety of approaches for classifying patients into a particular phenotype. Different techniques - natural language processing (NLP)-based techniques, rule-based systems, statistical analyses, data mining, or machine learning techniques - and data sources are used, and good performance is reported on datasets at respective institutions. However, no system makes comprehensive use of electronic medical records addressing all of their known weaknesses.

Holmes et al. [6] reviewed the literature on clinical research data warehouse governance in distributed research networks and reported interesting approaches regarding data stewardship, data privacy and security, query approval and user training, etc. Van Ommen et al. [17] present the recently established pan-European Biobanking and BioMolecular resources Research Infrastructure-European Research Infrastructure Consortium (BBMRI-ERIC) and the concept of Expert Centre as public-private partnerships in the precompetitive, not-for-profit field to provide a new structure to perform research projects that would face difficulties under currently established models of academic-industry collaboration. In this context, biological resources (cells, tissues, bodily fluids or biomolecules) are considered essential raw material for the advancement of health-related biotechnology, for research and development in life sciences, and for ultimately improving human health. Stored in local biobanks, access to the human biological samples and related medical data for transnational research is often limited, in particular for the international life science industry [17, 1].

Institutional Review Boards (IRBs) are also a critical component of clinical research and can become a significant bottleneck due to the dramatic increase, in both volume and complexity of clinical research [5]. A last challenge in the context of distributed research networks is to provide flexible and scalable service for record linkage that are widely needed to enable health researchers to mine longitudinal information for entire populations [2].

The last selected paper, by Rusanov et al. [11] focuses on data quality and the capability of the data to support research conclusions. The notion of data quality is complex, multi-dimensional and context dependent. The authors describe the relation between health status severity and the availability of data in EHR systems and demonstrate that it introduces a biasis in patient selection for clinical research.

In conclucion, current research demonstartes that different projects and initiatives could be well placed to deliver international solutions for the reuse of hospital EHR data to support clinical research studies. Selected articles demonstrate the potential of formal representation of multimodal and multi-level data in supporting interoperability across clinical research and care domains. With the development of pragmatic research, designed with input from health systems to support the identification of individual research subjects or cohorts as well as outcomes and produce evidence that can be readily used to improve care, a key issue for "learning health care organizations" is to systematically assess the quality of their data.

Acknowledgement

We would like to acknowledge the support of Martina Hutter and the reviewers in the selection process of the IMIA Yearbook.

References

1. Bowton E, Field JR, Wang S, Schildcrout JS, Van Driest SL, Delaney JT, et al. Biobanks and electronic medical records: enabling cost-effective research. Sci Transl Med 2014 Apr 30;6(234):234.
2. Boyd JH, Randall SM, Ferrante AM, Bauer JK, Brown AP, Semmens JB. Technical challenges of providing record linkage services for research. BMC Med Inform Decis Mak 2014 Mar 31;14:23.
3. Curtis JR, Wright NC, Xie F, Chen L, Zhang J, Saag KG, et al. Use of health plan combined with registry data to predict clinical trial recruitment. Clin Trials 2014 Feb;11(1):96-101.
4. Doods J, Botteri F, Dugas M, Fritz F; EHR4CR WP7. A European inventory of common electronic health record data elements for clinical trial feasibility. Trials 2014 Jan 10;15:18.
5. He S, Narus SP, Facelli JC, Lau LM, Botkin JR, Hurdle JF. A domain analysis model for eIRB systems: Addressing the weak link in clinical research informatics. J Biomed Inform 2014 Dec;52:121-9.
6. Holmes JH, Elliott TE, Brown JS, Raebel MA, Davidson A, Nelson AF, et al. Clinical research data warehouse governance for distributed research networks in the USA: a systematic review of the literature. J Am Med Inform Assoc 2014 Jul-Aug;21(4):730-6.
7. Jiang G, Evans J, Oniki TA, Coyle JF, Bain L, Huff SM, et al. Harmonization of detailed clinical models with clinical study data standards. Methods Inf Med 2015 Jan 12;54(1):65-74.
8. Kahn MG, Weng C. Clinical research informatics: a conceptual perspective. J Am Med Inform Assoc 2012 Jun;19(e1):e36-42.
9. Moor GD, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, et al. Using Electronic Health Records for Clinical Research: the Case of the EHR4CR Project. J Biomed Inform 2014 Oct 18.
10. Murphy SN, Herrick C, Wang Y, Wang TD, Sack D, Andriole KP, et al. High Throughput Tools to Access Images from Clinical Archives for Research. J Digit Imaging 2014 Oct 15.
11. Rusanov A, Weiskopf NG, Wang S, Weng C. Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research. BMC Med Inform Decis Mak 2014 Jun 11;14:51.
12. Sahoo SS, Lhatoo SD, Gupta DK, Cui L, Zhao M, Jayapandian C, et al. Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care. J Am Med Inform Assoc 2014 Jan-Feb;21(1):82-9.
13. Schreiweis B, Trinczek B, Köpcke F, Leusch T, Majeed RW, Wenk J, et al. Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials. Int J Med Inform 2014 Nov;83(11):860-8.
14. Shivade C, Raghavan P, Fosler-Lussier E, Embi PJ, Elhadad N, Johnson SB, et al. A review of approaches to identifying patient phenotype cohorts using electronic health records. J Am Med Inform Assoc 2014 Mar-Apr;21(2):221-30.
15. Tate AR, Beloff N, Al-Radwan B, Wickson J, Puri S, Williams T, et al. Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. J Am Med Inform Assoc 2014 Mar-Apr;21(2):292-8.
16. Trifirò G, Coloma PM, Rijnbeek PR, Romio S, Mosseveld B, Weibel D, et al. Combining multiple healthcare databases for postmarketing drug and vaccine safety surveillance: why and how? J Intern Med 2014 Jun;275(6):551-61.
17. van Ommen GJ, Törnwall O, Bréchot C, Dagher G, Galli J, Hveem K, et al. BBMRI-ERIC as a resource for pharmaceutical and life science industries: the development of biobank-based Expert Centres. Eur J Hum Genet 2014 Nov 19.
18. Xu W, Guan Z, Sun J, Wang Z, Geng Y. Development of an open metadata schema for prospective clinical research (openPCR) in China. Methods Inf Med 2014;53(1):39-46.

Correspondence to:
Christel Daniel, MD, PhD
INSERM UMRS 1142
CCS Patient – Assistance Publique – Hôpitaux de Paris
05 rue Santerre - 75 012 PARIS
France
Tel: +33 1 48 04 20 29
E-mail: christel.daniel@crc.jussieu.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2015, Section 'Clinical Research Informatics'

Sahoo SS, Lhatoo SD, Gupta DK, Cui L, Zhao M, Jayapandian C, Bozorgi A, Zhang GQ

Epilepsy and seizure ontology: towards an epilepsy informatics infrastructure for clinical research and patient care

J Am Med Inform Assoc 2014 Jan-Feb;21(1):82-9

Manual curation of heterogeneous data is a costly and non-scalable approach for studies involving thousands of patients. The objective of the paper is to demonstrate the benefit of using a formal ontology as common terminological resources to automatically reconcile data heterogeneity and implement large-scale, distributed data management systems. The authors created a multi-dimensional classification of epileptic seizures and epilepsy concepts - Epilepsy and Seizure Ontology (EpSO) - to consistently annotate epilespsy related information in order to share and integrate these data but also to create analytical software applications that can mine EpSO annotated data for knowledge discovery. EpSO, based on existing and widely used classification systems - a four-dimensional epilepsy classification system that integrates the latest International League Against Epilepsy terminology recommendations and National Institute of Neurological Disorders and Stroke (NINDS) common data elements - is a single domain model of more than 1000 classes with multiple levels of abstraction represented in OWL2[2]. The top-level classes in EpSO were defined as subclasses of the Basic Formal Ontology (BFO) upper-level ontology. Concepts from existing biomedical ontologies - such as the NeuroElectroMagnetic Ontologies (NEMO) and the Systematized Nomenclature of Medicine—Clinical Terms (SNOMED CT) - were imported in EpSO to model anatomy, drug information, electrophysiological data, and genetic information.

EpSO was integrated with a variety of informatics tools to i) streamline the entry of patient information, ii) process clinical free text in patient records and iii) run federated queries to create patient cohorts across multiple study centers. A new EpSO-based tool for epilepsy signal data analysis is also under development. This web-based signal visualization and analysis tool, called Cloudwave, maps event annotations to EpSO classes and uses it to query for specific segments of signal data. The Cloudwave signal processing module has been integrated with the open source Hadoop platform for distributed data processing and faster processing of signal segments.

De Moor G, Sundgren M, Kalra D, Schmidt A, Dugas M, Claerhout B, Karakoyun T, Ohmann C, Lastic PY, Ammour N, Kush R, Dupont D, Cuggia M, Daniel C, Thienpont G, Coorevits P

Using electronic health records for clinical research: the case of the EHR4CR project

J Biomed Inform 2015 Feb;53:162-73

Note: The paper was selected based on external reviews only since there was confilct of interest with one of the section editors (this review is written by the other section editor).

Enabling re-use of care produced data for research is now gaining interest in order to facilitate public health and research in the coming years.

EHR4CR is a IMI project that aims at offering Europe a large scale integrated infrastructure to help in identifying eligible patients for clinical research based on EHR data. The project is now within its last year (2011-2015) and the selected paper shows the current outcomes of the built infrastructure. The paper shows concretly how the plateform operates at EU level.

The paper presents the reference IT architecture defined to serve as a technical specification of the scalable plateform supporting EHR4CR research activities at EU level. It is composed of a set of common components and services facilitating the extraction of data from clinical systems as well as workflow interactions, privacy protection, information security and compliance with ethical, legal and regulatory local requirements.

The central EHR4CR architecture acts as a EU broker for clinical research. It has re-used software components from existing relevant projects such as i2b2 whenever possible. The development of the query builder interface was specified based on other succesful projects to compute eligibility criteria. A graphical query builder allows the user to combine core data elements (diagnoses, procedures, lab results and medications) with logical operators to represent free text eligibility criteria. The query builder developed also enables some temporal operators (Before, After, At most 3 months before, etc).

Each data endpoint then has to provide the necessary local query translation in order to be compatible with the central broker. EHR4CR promotes the use of international standards to operate, helping the local sites into leveradging their datasets for sharing in a wider context and for eventual future other regional, nation or international projects.

The EHR4CR platform, based on a Service Oriented Architecture (SOA), offers a set of services to operate smoothly : registry service for organisation-level metadata exposing services, Single Sign On (SSO) service for access autorisation, identity management service, message broker, terminology service, terminology mapping service, query service, patient recruitement and participation services.

The overall infrastructure was tested across several EU participating hospitals (11 hospitals). Not all EHR data elements required to answer EFPIA partners eligibility criterias were present in local information systems. Over the 375 free-text eligibility criteria proposed and reviewed, 175 were transformed into a computable representation and tested. Pilot sites mapped over 300 codes from their local terminologies into the central EHR4CR terminology. Variations in coding were observed. The technical platform was positively evaluated also in terms of usability.

[2] EpSO is available for download and use under a Creative Commons license at: http://prism.case.edu/prism/index.php/EpilepsyOntology

The EHR4CR infrastructure is well placed to deliver a sound, operational solution to accelerate the building of european clinical research networks. The platform integrated all required components to be widely used at a larger scale. The required cost to integrate the overall network for an hospital should be evaluated in order to measure not its technical mean to meet its objectives, but the local effort required to participate to the EHR4CR network.

Schreiweis B, Trinczek B, Köpcke F, Leusch T, Majeed RW, Wenk J, Bergh B, Ohmann C, Röhrig R, Dugas M, Prokosch HU

Comparison of electronic health record system functionalities to support the patient recruitment process in clinical trials

Int J Med Inform 2014 Nov;83(11):860-8

Most projects conducted to allow reusing EHR data for patient recruitment focus on standalone systems and proprietary implementations at one particular institution often for only one singular trial and are typically limited to few research institutions with IT development teams at site.

The authors selected five sites from a portfolio of German University Hospitals runing EHR implementations and evaluated within these sites the existence of modules or tools, which can readily be applied for IT-supported patient recruitment scenarios. Electronic support for patient recruitment process requires five specific modules which are: data storage (either theEHR database or the data warehouse), a local trial registry, a query module, a notification tool and ascreening list module.

At the five selected sites, with four commercial EHR systems and one self-developed system, the authors found that – even though no dedicated module for patient recruitment has been provided –EHR products comprise generic tools such as workflow engines, querying capabilities, report generators and direct SQL-based database access which can be applied as query modules, screening lists and notification components for patient recruitment support. However, the screening list module in all systems could currently only be implemented as workaround solutions and a major limitation of all current EHR products is that they provide no dedicated data structures and functionalities for implementing and maintaining a local trial registry.

Rusanov A, Weiskopf NG, Wang S, Weng C

Hidden in plain sight: bias towards sick patients when sampling patients with sufficient electronic health record data for research

BMC Med Inform Decis Mak 2014 Jun 11;14:51

While the re-use of electronic clinical data is forseen as a potential future for public health and patient recruitment for clinical trials, caveats are existing and are presented in this paper. As often observed, EHR data produced into daily clinical practice suffers from data quality problems. The most common data quality problem is the data sufficiency (completeness). Even when the data can and should be recorded within the clinical task, it could be under-documented for various reasons (not clinically relevant to document, could not be performed, etc.). As observed, missing data is very common in today's EHR databases. The authors present here a study to show the correlation between a patient's health status and the amount of available data to characterize its health status or phenotype. This may cause a major biasis of the possible outcomes of the research since the study would be processed on sicker patients, which could have a dramatic effect on the generated knowledge.

The authors present their methodology. They first found an adapted classification of patient status from the American Society of Anesthesiologists which represents 6 health status inside which patients can be classifed. They queried their anesteshia research database of all cases having an ICD-9 code. The data was then compared to their local Clinical Data Warehouse to obtain the number of days with medication orders and the number of days with laboratory results for each patient. All data were standardised.

The results of the negative binomial regression model based on 10,000 cases showed that the relation between patient health status and EHR data sufficiency is significant. This proves that when limiting exploitation of EHRs to complete or sufficient data sets per given patient induces a biasis in representativity of the studied population. The authors recognize their study should be tested at other care facilities to validate its scalability. The authors finally alert those who are using EHR data for secondary use they should be aware of such biasis.