Big Data and Smart Health Strategies: Findings from the Health Information Systems Perspective

L. Toubiana[1], M. Cuggia[2], Section Editors for the IMIA Yearbook Section on Health Information Systems

[1] INSERM, U1142, LIMICS, F-75006, Paris, France; Sorbonne Universités, UPMC Univ Paris 06, UMR_S 1142, LIMICS, F-75006, Paris, France; Université Paris 13, Sorbonne Paris Cité, LIMICS, (UMR_S 1142), F-93430, Villetaneuse, France.

[2] INSERM U936, Faculté de Médecine, DIM - CHU Pontchaillou, Rennes, F-35000, France

Summary

Objectives: To summarize excellent current research in the field of Health Information Systems.

Method: Creation of a synopsis of the articles selected for the 2014 edition of the IMIA Yearbook.

Results: Four papers from international peer reviewed journals were selected and are summarized.

Conclusions: Selected articles illustrate current research regarding the impact and the evaluation of health information technology and the latest developments in health information exchange.

Keywords

Medical informatics, International Medical Informatics Association, health information systems, hospital information systems

Introduction

One objective of the IMIA Yearbook is to review research articles published in the preceding year to identify research trends and excellent research papers in the different fields of medical informatics that reflect a broad spectrum of topics. The Health information systems (HIS) section has its own wide spectrum of research topics: methodical aspects, analysis, design, evolution, and evaluation of HIS, as well as technology assessment.

The aim of health information systems is to contribute to the high quality of care and the efficiency of care management, for physicians, patients, and medical research. With the growing deployment of health information technology (HIT), the question of effectiveness is still topical, as reflected by this year's theme for the Yearbook 'Big Data and Smart Health Strategies'.

About the Paper Selection

As in previous years, the best papers selection for the HIS section of the 2014 IMIA Yearbook represents examples of excellent research in the field of health information systems. A comprehensive literature search was performed, using two bibliographic databases, Pubmed/Medline (from NCBI, National Center for Biotechnology Information) and Web of Science® (from Thomson Reuters). The search, targeted to HIS, yield-ed a total of 3, 482 references from which a set of 209 references from 28 international peer reviewed journals was pre-selected. These references were blindly reviewed by the two section editors and classified in three categories: kept, discarded, or pending. In this review process, papers were evaluated according to their contribution to one of the following themes: originality of the application domain, description of a new method, evaluation, and outcomes. Then, reviews from both section editors were merged, yielding 22 references accepted by both reviewers (with the status kept – kept). These 22 papers were reviewed jointly by the two section editors to reach a consensual list of 15 candidate best papers. Following the IMIA Yearbook process, these 15 papers where then evaluated by the Yearbook editors and external reviewers (at least four reviewers per paper).

Four papers were finally selected from international peer reviewed journals whose topic areas were in the fields of medicine and medical informatics. Table 1 lists the selected papers. A content summary of the selected papers can be found in the appendix of this synopsis.

Conclusion and Outlook

As in previous years, the selection of best papers from a large corpus of published studies was performed according to quality criteria, independent of the special theme of

2014, which is "Big Data and Smart Health Strategies". In 2013, the selection performed using the same criteria had highlighted the multidimensional aspects involved in the deployment of health information systems. In contrast, this year's selection gives the impression of a much tighter focus around the issue of data sharing among multiple centers.

Thus, the first paper presented in this selection describes a method to extract data in the context of pharmacovigilance from eight European EHR databases. The study data was comprised of a follow-up of almost 60 million person-years. An iterative harmonization process enabled a more homogeneous identification of events across these differently structured databases. In the second selected paper, authors propose a method that provides simple identification of pre-research cohorts merged with public health data sources in the USA. The platform provides a front-end graphical user interface and enables researchers to build and run queries, and access an aggregated visualization of results. This architecture uses well-described technologies that should be possible to implement for most institutions. The third paper describes an implementation of a self-scaling registry for collaborative data sharing based upon the Integrating Biology & the Bedside (i2b2) data warehousing framework and SHRINE, an open source software, to create a modular, ontology-based, federated infrastructure. This approach can be extended as a scalable, reusable solution facilitating interdisciplinary research across diseases. The fourth paper reports a solution for facilitating semantic interoperability to rule-based Clinical Decision Support Systems (CDSSs) focusing on standardized input and output documents conforming an HL7-CDA wrapper.

Unintentionally, three of the four best papers selected by the reviewers were published in the same journal. This reinforces the idea that the issue of large scale data sharing, and the selection of best papers for this year responded spontaneously to the special theme of the year. This clearly demonstrates that HIS projects and implementations addressed the issue of providing global knowledge and Smart Health Strategies from huge amount of data collected locally.

Table 1 Best paper selection of articles for the IMIA Yearbook of Medical Informatics 2014 in the section 'Health Information Systems'. The articles are listed in alphabetical order of the first author's surname.

Section
Health Information Systems
▪ Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, Mazzaglia G, Giaquinto C, Fornari C, Herings R, Molokhia M, Pedersen L, Fourrier-Réglat A, Fieschi M, Sturkenboom M, van der Lei J, Pariente A, Trifirò G; EU-ADR consortium. Harmonization Process for the Identification of Medical Events in Eight European Healthcare Databases: The Experience from the EU-ADR Project. J Am Med Inform Assoc 2013; 20(1):184-92.
▪ Hurdle JF, Haroldsen SC, Hammer A, Spigle C, Fraser AM, Mineau GP, Courdy SJ. Identifying Clinical/translational Research Cohorts: Ascertainment via Querying an Integrated Multi-Source Database. J Am Med Inform Assoc 2013; 20(1):164-71.
▪ Natter MD, Quan J, Ortiz DM, Bousvaros A, Ilowite NT, Inman CJ, Marsolo K, McMurry AJ, Sandborg CI, Schanberg LE, Wallace CA, Warren RW, Weber GM, Mandl KD. An i2b2-Based, Generalizable, Open Source, Self-Scaling Chronic Disease Registry. J Am Med Inform Assoc 2013; 20(1): 172-9.
▪ Sáez C, Bresó A, Vicente J, Robles M, García-Gómez JM . An HL7-CDA Wrapper for Facilitating Semantic Interoperability to Rule-Based Clinical Decision Support Systems. Comput Methods Programs Biomed 2013;109(3): 239-49.

Correspondence to:
Laurent Toubiana, PhD
INSERM UMRS 1142 "LIMICS"
15, rue de l'École de Médecine
75006 Paris, France
Tel: +33 1 44 27 91 97
E-mail: Laurent.toubiana@inserm.fr

Appendix: Content Summaries of Selected Best Papers for the IMIA Yearbook 2014, Section 'Health Information Systems'

Avillach P, Coloma PM, Gini R, Schuemie M, Mougin F, Dufour JC, Mazzaglia G, Giaquinto C, Fornari C, Herings R, Molokhia M, Pedersen L, Fourrier-Réglat A, Fieschi M, Sturkenboom M, van der Lei J, Pariente A, Trifirò G; EU-ADR consortium

Harmonization Process for the Identification of Medical Events in Eight European Healthcare Databases: The Experience from the EU-ADR Project

J Am Med Inform Assoc 2013; 20(1):184-92

This paper describes the procedure used for harmonizing the extraction from eight European EHR databases for five events deemed to be important in pharmacovigilance. Clinical information is collected using four different disease terminologies and free text in two different languages. A common database model was used to share and pool data and verify the semantic basis of the event extraction queries. The study population comprised 19,647,445 individuals with a follow-up of 59,929,690 person-years. Standardized and age-specific incidence rates were calculated to facilitate benchmarking and harmonization of event data extraction across the databases. The iterative harmonization process enabled a more homogeneous identification of events across differently structured databases using different coding-based algorithms.

Hurdle JF, Haroldsen SC, Hammer A, Spigle C, Fraser AM, Mineau GP, Courdy SJ

Identifying Clinical/translational Research Cohorts: Ascertainment via Querying an Integrated Multi-Source Database

J Am Med Inform Assoc 2013; 20(1):164-71

This paper proposes a method to use data from merged public health data sources, available in most states of the USA, to identify pre-research cohorts. The approach incorporates several coding systems for medical information. It provides a front-end graphical user interface and enables

researchers to build and run queries and view aggregated results. The tool has been rapidly adopted: 241 users representing a wide range of disciplines from 17 institutions have signed the user agreement and used the query tool. This architecture uses well-described technologies that should be possible to implement for most institutions.

Natter MD, Quan J, Ortiz DM, Bousvaros A, Ilowite NT, Inman CJ, Marsolo K, McMurry AJ, Sandborg CI, Schanberg LE, Wallace CA, Warren RW, Weber GM, Mandl KD

An i2b2-Based, Generalizable, Open Source, Self-Scaling Chronic Disease Registry

J Am Med Inform Assoc 2013; 20(1): 172-9

This paper describes the implementation of a self-scaling registry for collaborative data sharing based upon the Integrating Biology & the Bedside (i2b2) data warehousing framework and SHRINE, an open source software, to create a modular, ontology-based, federated infrastructure. The 56-site Childhood Arthritis & Rheumatology Research Alliance (CARRA) Registry and 3-site Harvard Inflammatory Bowel Diseases Longitudinal Data Repository now utilize this technology. This infrastructure allows researchers to share and combine their research datasets. The implementation of i2b2-SSR for multi-sites has established a digital infrastructure for community-driven research data sharing in pediatric rheumatology in the USA. This approach can be extended as a scalable, reusable solution facilitating interdisciplinary research across diseases.

Sáez C, Bresó A, Vicente J, Robles M, García-Gómez JM

An HL7-CDA Wrapper for Facilitating Semantic Interoperability to Rule-Based Clinical Decision Support Systems

Comput Methods Programs Biomed 2013;109(3): 239-49

The authors of the paper propose a solution for facilitating semantic interoperability in rule-based Clinical Decision Support Systems (CDSSs) focusing on standardized input and output documents conforming to HL7-CDA wrapper. Patient data and rule inference results are mapped respectively to and from the CDSS by means of a binding method based on an XML binding file. The proposed solution is being applied in a CDSS for providing patient-specific recommendations for the care management of outpatients with diabetes mellitus.