Electronic Patient Records: Some Answers to the Data Representation and Reuse Challenges

Findings from the Section on Patient Records

S. Meystre, Managing Editor for the IMIA Yearbook Section on Patient Records
University of Utah, Department of Biomedical Informatics, Salt Lake City, UT, USA

Summary

Objectives: To summarize current excellent research in the field of patient records.

Method: Synopsis of the papers selected for the IMIA Yearbook 2007.

Results: The Electronic Patient Record encompasses a broad field of research and development. Some current research topics were selected for this IMIA Yearbook: EHR representation and communication standards, and secondary uses of clinical data for research and decision support. Four excellent papers representing the research in those fields were selected for the Patient Records section.

Conclusion: The best papers selected for this section focus on the analysis and comparison of two important clinical documents representation standards, on direct structured data entry, on the use of Natural Language Processing to detect adverse events, and on the development and evaluation of a clinical text corpus annotated for part-of-speech information.

Keywords

Medical informatics, International Medical Informatics Association, yearbook, medical records, computerized medical record systems, natural language processing

Geissbuhler A, Haux R, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2007. Methods Inf Med 2007; 46 Suppl 1: 47-9

Introduction

An Electronic Patient Record (EPR) may simply be seen as a shared repository to record and view observations, decisions, and intended actions relating to a patient, but it also combines several difficult requirements such as mobile patients, data shared and authored by multiple users simultaneously, and wide geographical availability of a given record to multiple carers and applications [1]. The EPR is often called Electronic Health Record (EHR) [2] when providing an integrated, longitudinal, cross-institutional record. The latter requires information modeling and communication standardization to realize interoperability. To this end, several EHR standards are currently under development [3] and some are discussed in one of the best papers selected for this section [4].

Other important trends of research and development in the EPR domain include the integration of other types of data such as genomic data [5] and also include means to ease data reuse for secondary purposes such as research and decision support. Structured data entry is a possible solution to allow reusing clinical data for research and is investigated in a selected best paper [6]. Natural Language Processing (NLP) to extract structured data from narrative text is another solution and is investigated in two other selected best papers [7,8].

Best Paper Selection

Selection of the best papers for the Patient Records section in the IMIA Yearbook 2007 reflects the research and development trends cited above and also already expressed in a review [9] and in previous editions of the IMIA Yearbook [10,11].

As a result of a comprehensive review process based on criteria cited in [12], four outstanding papers representing the research in the EPR domain were selected from peer-reviewed journals in the field of medical informatics. Table 1 lists the selected papers. A brief content summary of these selected papers can be found in the Appendix of this synopsis.

Conclusions and Outlook

As cited above, EHR standards are required to share health record data whilst preserving faithfully the original meaning intended by the author [3]. Important standards focusing on the representation and exchange of clinical documents are the HL7 Clinical Document Architecture (CDA) [13] and the ASTM Continuity of Care Record (CCR) [14]. These standards are analogous and [4] analyzes this similarity and proposes some solutions to improve their interoperability. An agreement between HL7 and the ASTM has resulted in the development of the Continuity of Care

Table 1 Best paper selection for the IMIA Yearbook of Medical Informatics 2007 in the section 'Patient Records'. The articles are listed in alphabetical order of the first author's surname.

Section
Patient Records
▪ Ferranti JM, Musser RC, Kawamoto K, Hammond WE. The clinical document architecture and the continuity of care record: a critical analysis. J Am Med Inform Assoc 2006 May-Jun;13(3):245-52.
▪ Los RK, van Ginneken AM, van der Lei J. OpenSDE: a strategy for expressive and flexible structured data entry. Int J Med Inform 2005 Jul;74(6):481-90.
▪ Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc 2005 Jul-Aug;12(4):448-57.
▪ Pakhomov SV, Coden A, Chute CG. Developing a corpus of clinical notes manually annotated for part-of-speech. Int J Med Inform 2006 Jun;75(6):418-29

Document (CCD). The CCD maps CCR elements into a CDA representation and its final version has been approved by HL7 in January 2007.

Direct data entry remains an important challenge and requires an easy, flexible and rapid interface, and above all, no need to enter the same information twice (e.g. once for clinical use and once for research). Structured data entry has several advantages such as a more complete documentation of care and the possible reuse of clinical data for research. Documenting the physical examination seems adapted to structured data entry, but less structured types of documentation such as the patient history require a more flexible and expressive format: usually narrative text [6]. The patient record contains a considerable amount of information, but the recorded clinical information is commonly unstructured narrative text. These narrative text documents make up a substantial part of the medical record, providing information leading to the final diagnosis in 76% of the cases [15]. Narrative text is still the most user-friendly and expressive way of recording information, but the increasing use of encoded data and the requirement for standard medical data set creates a need for coded information instead. As a possible response to this problem, NLP can convert narrative text into structured data, and therefore extend the use of the EPR [16]. The application of NLP to detect adverse events from narrative text is demonstrated in [7].

Annotated clinical text corpora are still very rare and required in the biomedical domain to effectively train and evaluate NLP applications that are based on statistical methods. Pakhomov et al. [8] describe the development and evaluation of a clinical text corpus annotated for part-of-speech information. Up-to-date information about current and future issues of the IMIA Yearbook is available at http://www.schattauer.de/index.php?id=1384

Acknowledgments

We greatly acknowledge the support of Martina Hutter and of the reviewers in the papers selection process of the IMIA Yearbook.

References

1. Beale T. The Health Record - why is it so hard? In: Haux R, Kulikowski C, editors. IMIA Yearbook of Medical Informatics 2005. Stuttgart: Schattauer; 2004. p. 301-4.
2. Nøhr C. Evaluation of Electronic Health Record Systems. Methods Inf Med 2006;45 Suppl 1:107-13.
3. Kalra D. Electronic Health Record Standards. Methods Inf Med 2006;45 Suppl 1:136-44.
4. Ferranti JM Musser RC, Kawamoto, K, Hammond WE. The clinical document architecture and the continuity of care record: a critical analysis. J Am Med Inform Assoc 2006 May-Jun;13(3):245-52.
5. Sax U, Schmidt S. Integration of genomic data in Electronic Health Records - opportunities and dilemmas. Methods Inf Med. 2005;44(4):546-50.
6. Los RK, van Ginneken AM, van der Lei J. OpenSDE: a strategy for expressive and flexible structured data entry. Int J Med Inform 2005 Jul;74(6):481-90.
7. Melton GB, Hripcsak G. Automated detection of adverse events using natural language processing of discharge summaries. J Am Med Inform Assoc 2005 Jul-Aug;12(4):448-57.
8. Pakhomov SV, Coden A, Chute CG. Developing a corpus of clinical notes manually annotated for part-of-speech. Int J Med Inform 2006 Jun;75 (6):418-29.
9. Jaspers MWM, Knaup P, Schmidt D. The Computerized Patient Record: Where Do We Stand? Methods Inf Med. 2006;45 Suppl 1:29-39.
10. Knaup P. Electronic Patient Records and their benefit for Patient Care. Methods Inf Med 2006;45 Suppl 1:40-2.
11. Bott OJ, Ammenwerth E, Brigl B, Knaup P, Lang E, Pilgram R, et al. The challenge of ubiquitous computing in health care: technology, concepts and solutions. Findings from the IMIA Yearbook of Medical Informatics 2005. Methods Inf Med 2005;44(3):473-9.
12. Ammenwerth E, Wolff AC, Knaup P, Ulmer H, Skonetzki S, van Bemmel J, et al. Developing and Evaluating Criteria to Help Reviewers of Biomedical Informatics Manuscripts. J Am Med Inform Assoc 2003;10:512-14.
13. Dolin RH, Alschuler L, Boyer SL, Beebe C, Behlen FM, Biron PV, et al. HL7 Clinical Document Architecture, Release 2. J Am Med Inform Assoc 2006;13:30-9.
14. Kibbe DC, Phillips RL, Green LA. The Continuity of Care Record. Am Fam Physician 2004; 70:1220-3.
15. Peterson MC, Holbrook JH, Von Hales D, Smith NL, Staker LV. Contributions of the history, physical examination, and laboratory investigation in making medical diagnoses. West J Med 1992 Feb;156(2):163-5.
16. Spyns P. Natural language processing in medicine: an overview. Methods Inf Med 1996 Dec;35(4-5):285-301.
17. OpenSDE. Available from: http://sourceforge.net/projects/opensde
18. Van der Lei J. Closing the loop between clinical practice, research and education: the potential of electronic patient records. Methods Inf Med 2002;41(1):51-4.
19. Roukema J, Los RK, Bleeker SE, van Ginneken AM, van der Lei J, Moll HA. Paper versus computer: feasibility of an electronic medical record in general pediatrics. Pediatrics 2006 Jan; 117(1):15-21.
20. Los RK, Roukema J, Van Ginneken AM, De Wilde M, Van der Lei J. Are Structured Data Structured Identically? Investigating the Uniformity of Pediatric Patient Data Recorded Using OpenSDE. Methods Inf Med. 2005;44:631-8.
21. Friedman C, Alderson PO, Austin JH, Cimino JJ, Johnson SB. A general natural-language text processor for clinical radiology. J Am Med Inform Assoc 1994 Mar-Apr;1(2):161-74.
22. Marcus M, Santorini B, Marcinkiewicz MA. Building a large annotated corpus of English: the Penn Treebank. Linguistics 1993;19:297-352.
23. Brants T. TnT - a statistical part-of-speech tagger. NAACL/ANLP; 2000.

Correspondence to:
Stéphane Meystre
University of Utah
Department of Biomedical Informatics
Salt Lake City, Utah, USA
s.meystre@utah.edu

Appendix: Content Summaries of Selected Best Papers, Patient Records Section*

Ferranti JM, Musser RC, Kawamoto K, Hammond WE

The clinical document architecture and the continuity of care record: a critical analysis

J Am Med Inform Assoc 2006 May-Jun; 13(3):245-52

Several Standards Development Organizations (SDO) are currently working on frameworks for representing and exchanging the content of EHRs. Prominent examples of those frameworks are the HL7 Clinical Document Architecture (CDA) [13] and the ASTM International Continuity of Care Record (CCR) [14]. Release 2 of the former has been published in 2005 and derives its content from the HL7 Reference Information Model. It was conceived to represent virtually any type of clinical document. Version 1.0 of the CCR was also published in 2005 and is mainly intended to provide consulting physicians with the information necessary to participate in a patient's care. It is focused on the primary care "summary record".
These standards are unfortunately similar. The authors see substantial value in the CCR but are concerned that it does not possess the same potential for interoperability as the CDA. They compare both standards and propose several interoperability solutions.

Los RK, van Ginneken AM, van der Lei J

OpenSDE: a strategy for expressive and flexible structured data entry

Int J Med Inform 2005 Jul;74(6):481-90

This paper proposes a detailed description of a novel application called OpenSDE and allowing structured data entry. OpenSDE [17] is an open source application supporting data entry by clinicians in a variety of settings, for both patient care and research [18]. It is based on the selection of predefined concepts organized as nodes in a tree structure. Each node has a set of data items to specify such as its status, value, and a timestamp. Free text comments can also be added. The tree holds constraints and is domain specific. Trees are called domain models and are manually authored. They resemble terminologies and were developed instead of using standard terminologies because the latter have a different purpose and insufficient granularity and flexibility.
OpenSDE is used at the Erasmus Medical Center (Rotterdam, Netherlands) in several departments and has been mostly evaluated in Pediatrics [19,20].

Melton GB, Hripcsak G

Automated detection of adverse events using natural language processing of discharge summaries

J Am Med Inform Assoc 2005 Jul-Aug; 12(4):448-57

The authors have adapted a well-known natural language processing application called MedLEE [21] to detect adverse events in narrative text. The 45 patient-related hospital-based adverse event types defined in the New York Patient Occurrence Reporting and Tracking System (NYPORTS) were searched in discharge summaries of inpatients at the New York-Presbyterian Hospital-Columbia University Medical Center.
Agreement between reviewers was excellent (kappa = 0.94). The system was mainly evaluated with 1000 randomly selected cases from a set of 57452 inpatients in the years 1996 to 2000. A manual chart review of discharge summaries, of the full electronic chart, and of the paper chart, served as gold standard. Discharge summaries were shown to contain most of the information needed to detect NYPORTS adverse events. When aggregated by case, sensitivity of the system was 0.28 and specificity was 0.98. When aggregated by event, sensitivity was 0.25 and specificity was 0.99. Evaluation based on the full cohort of 57452 patients with review of the cases with adverse events detected by the system gave a positive predictive value of 0.45 when aggregated by case, and 0.44 when aggregated by event. These results compare favorably with other similar studies. The system was also compared with traditional adverse event reporting and exhibited a sensitivity of 0.34 when traditional methods had a sensitivity of only 0.086.

Pakhomov SV, Coden A, Chute CG

Developing a corpus of clinical notes manually annotated for part-of-speech

Int J Med Inform 2006 Jun;75(6):418-29

This paper describes the development and evaluation of a corpus of clinical text manually annotated for part-of-speech (POS) information. A corpus of 273 clinical notes form the Mayo Clinic were annotated with POS tags from the Penn Treebank tagset [22]. Annotation was done by three medical coding experts. Inter-annotator agreement was excellent with a kappa coefficient of 0.93. Using a variation of the 10-fold cross-validation, the authors trained the TnT tagger [23] on Penn Treebank data and on clinical data from the corpus, and then evaluated it with part of the annotated corpus. When only trained on Penn Treebank data, POS tagging correctness was 89.8%, and when also trained on clinical data, correctness grew to 94.7%, suggesting a need to adapt POS taggers to the clinical sublanguage. Evaluation of correctness for 10 different sections of clinical notes (e.g. current medications, allergies, family history) showed results between 74.7% for the current medications section and 92.6% for the family history section, suggesting that further adaptation to particular clinical note sections may improve the correctness of POS tagging.

* The complete papers can be accessed in the Yearbook's full electronic version, provided that permission has been granted by the copyright holder(s)