International Journal of Medical Informatics (2005) **74**, 387—394

ELSEVIER

International Journal of Medical Informatics

www.intl.elsevierhealth.com/journals/ijmi

Consequences of impaired data quality on information retrieval in electronic patient records

Gustav Mikkelsen[a,*], **Jan Aasly**[b,1]

[a] *Department of Clinical Neurosciences, The Norwegian University of Science and Technology, St. Olavs Hospital, 7006 Trondheim, Norway*
[b] *Department of Neurology, St. Olavs Hospital HF, Olav Kyrresgt. 17, 7006 Trondheim, Norway*

Received 16 December 2003; received in revised form 20 September 2004; accepted 3 November 2004

KEYWORDS
Evaluation studies;
Medical records;
Computerized;
Data quality;
Information storage and retrieval;
Neurology

Summary

Objectives: To assess the quality of specific information in electronic patient records and the consequences of sub-optimal data quality on automated information retrieval.
Methods: Patient records were evaluated with respect to accuracy of data relevant for retrieval according to a source-oriented, time-oriented and concept-oriented view of the record. Retrieval effectiveness was estimated using various methods based on record structure, text based retrieval and combinations of these.
Results: 98.1% of record documents were consistent regarding author, 99.8% regarding department of origin and 90.9% regarding document date. Document type was definitely not consistent in 8% of the documents. Estimated recall was 97% with 50% precision for document retrieval on the basis of date, and varying from 31 to 100% for retrieval based on document type. Retrieval based on manually supplied semantic tags performed better than simple string-based methods and improved when combined with string-matching mechanisms.
Conclusions: Data attributes central for automated document retrieval in electronic patient records showed variable accuracy, with potentially negative consequences for basic record navigation. Text-based retrieval was inferior to methods based on data representing record structure. Quality of specific information elements suffered from lack of precise definitions and adequate mechanisms for quality assurance.

1. Introduction

One potential benefit of electronic patient record (EPR) systems is to make record information more readily available for legitimate users [1]. By applying automated retrieval mechanisms, specific information may be efficiently retrieved from large

* Corresponding author. Tel.: +47 73 86 84 20; fax: +47 73 86 75 81.
E-mail addresses: gustav.mikkelsen@stolav.no (G. Mikkelsen), jan.aasly@medisin.ntnu.no (J. Aasly).
[1] Tel.: +47 73 86 84 14; fax: +47 73 86 75 81.

1386-5056/$ — see front matter © 2004 Elsevier Ireland Ltd. All rights reserved.
doi:10.1016/j.ijmedinf.2004.11.001

amounts of diverse patient record information [2—4]. Although, progress in information technology has created new opportunities with respect to handling large volumes of complex information, performance of mechanisms for automated retrieval still depend on the quality of the underlying data. Incorrect or missing information will affect the retrieval process adversely [4—7], as relevant information may be missed and irrelevant information may be retrieved.

The purposes of the medical record are numerous and users of the record have different foci and needs. Correspondingly, the information in the patient record is diverse, voluminous and of variable quality [8—13]. Different kinds of information in the records have variable relevance for different kinds of users. Data considered of little importance by some users may be important to others. Data entered into an EPR system are more likely to be of high quality when the user responsible for data entry considers the information to be important [5,9—11,14—16]. Correspondingly, information considered less important by those responsible for recording, it is more likely to be of sub-optimal quality [17].

Basic mechanisms for retrieving and presenting information in electronic patient record systems may include mechanisms for navigating the record according to when the information was produced, who or what produced the information and how the information has been categorized according clinical relevance. For example, when navigating to a specific discharge summary in a document based system, information about who produced it, when it was produced and the fact that the document has been categorized as a ''Discharge summary'' may all be exploited to locate the document. This way of retrieving information depends on how information in the record system is organized and is frequently based on other data than those in focus by ordinary users of the system. The purpose of some information in the patient record is to identify and characterize the medically relevant information in the patient record. This information, representing the structure of the record, may be provided automatically by the system, e.g. patient identifiers, timestamps etc., or the users of the system may enter them manually into the system. Clinical personnel may regard this supplementary information of modest value, as it has no direct significance for tasks related to patient care. The role for system operation may not always be evident to clinical users, especially if the EPR system is complex and users are marginally trained. Subsequently, the efforts spent on ensuring high quality of information representing the structure of the record may have variable effects [6,7,18].

Structuring of the medical record has been debated extensively for decades. Different views exist regarding how patient records should be organized in order to function optimally for clinical users, researchers, administrators and other legitimate users of the patient records [19—24]. Traditional views of the record are: (1) source-oriented views, i.e. based upon who or what has produced the information, (2) time-oriented views, i.e. based upon the time the information was produced or the events took place and (3) concept-oriented or problem-oriented views, i.e. based upon the relevance of specific information to various medical concepts or clinical problems [3,25]. Retrieving or presenting information according to these axes usually depends on specific data stored in the electronic patient records. Missing or incorrect association of record elements with identifiers for source, time or specific clinical concepts are likely to limit the possibility of retrieving record information according to these aspects.

Given this background, we wished to evaluate the accuracy, i.e. the completeness and correctness [13] of specific information in electronic patient records and estimate how efficient information could be retrieved, based upon the views of the record described above. Correctness of data used for basic record navigation and presentation according to source, time and clinical concepts, was assessed in a document-based electronic patient record system. Effectiveness of document retrieval on basis of central document attributes (author, department, type and date) and manually supplied indexes identifying clinical concepts (semantic tags) was estimated as recall and precision [26]. Methods for retrieval based on simple string matching were evaluated as reference.

2. Material and methods

2.1. Setting

The study was performed in a neurological department in a Norwegian university hospital. An EPR system for handling textual record notes is used in parallel with the paper-based patient records. The patient record is shared among several departments. Record notes dictated by doctors are typed into the electronic system by secretaries and printed to the paper record. In this way, both electronic and hardcopy versions are obtained for most textual record documents. All notes contain