World Academy of Science, Engineering and Technology
International Journal of Health and Medical Engineering
Vol:1, No:11, 2007

The Relevance of Data Warehousing and Data Mining in the Field of Evidence-based Medicine to Support Healthcare Decision Making

Nevena Stolba and A Min Tjoa

Abstract**—** Evidence-based medicine is a new direction in modern healthcare. Its task is to prevent, diagnose and medicate diseases using medical evidence. Medical data about a large patient population is analyzed to perform healthcare management and medical research. In order to obtain the best evidence for a given disease, external clinical expertise as well as internal clinical experience must be available to the healthcare practitioners at right time and in the right manner. External evidence-based knowledge can not be applied directly to the patient without adjusting it to the patient's health condition. We propose a data warehouse based approach as a suitable solution for the integration of external evidence-based data sources into the existing clinical information system and data mining techniques for finding appropriate therapy for a given patient and a given disease. Through integration of data warehousing, OLAP and data mining techniques in the healthcare area, an easy to use decision support platform, which supports decision making process of care givers and clinical managers, is built. We present three case studies, which show, that a clinical data warehouse that facilitates evidence-based medicine is a reliable, powerful and user-friendly platform for strategic decision making, which has a great relevance for the practice and acceptance of evidence-based medicine.

Keywords**—**data mining, data warehousing, decision-support systems, evidence-based medicine.

I. Introduction

EVIDENCE based medicine is the conscientious, explicit, and judicious use of current best evidence in making decisions about the care of individual patients [1]. The task of the evidence-based medicine is to complement the existing clinical decision making process with the most accurate and most efficient research evidence. For example, when treating a diabetes patient suffering from a progressive liver disease, his (her) clinician has to find the most efficient therapy, which does not conflict with the patient's ongoing diabetes treatment. The clinician searches through evidence-based guidelines to find current best evidence for treating liver diseases and verifies if the proposed method fits into the diabetes patient's health risks.

The term data warehouse was coined with the definition of Inmon: "A warehouse is a subject-oriented, integrated, time-variant and non-volatile collection of data in support of management's decision making process" [2].

Data mining has been defined as the nontrivial extraction of implicit, previously unknown, and potentially useful information from data [5].

The combination of data warehousing and data mining technology and evidence-based medicine as a new direction in modern health care commences an innovative application field of information technology in health care industry. Medical institutions as well as health insurance companies have primarily interest in increasing the patient healing rate and reducing treatment costs. Application of data warehousing in the area of evidence-based medicine could prove economical in the long term by avoiding the duplication of examinations, time saved through automation of routine tasks and the simplification of accounting and administrative procedures.

The goal of this paper is to show how data warehousing supports the use of evidence-based medicine. We show the benefits, data warehouse is bringing in the area of creating evidence-based rules, delivering evidence-based guidelines and alerts at the point of patient care and to support clinical or business strategies and providing standing or ad-hoc reports for care givers (clinicians, physicians, nurses) and other decision-makers (clinical management, human resources).

The contribution of our paper is that to demonstrate that the data warehouse approach proposed constitutes a reliable and efficient carrier of the clinical decision support system facilitating evidence-based medicine.

In chapter 2, we briefly introduce the most relevant data sources for clinical decision support systems and we describe the most important application fields for decision support systems in the area of evidence-based medicine. Three case studies, which demonstrate the contribution of the utilisation of data warehousing and data mining in the clinical practise are given in chapter 3. Related work is presented in chapter 4, followed by the conclusion in chapter 5.

Manuscript received December 24, 2005. This research has been partly funded by the Austrian Federal Ministry for Education, Science and Culture, and the European Social Fund (ESF) under grant 31.963/46-VII/9/2002.

Nevena Stolba is with the Woman's Postgraduate College for Internet Technologies, Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria (phone: +431-58801-18818; fax: +431-58801-18895; e-mail: stolba@wit.tuwien.ac.at).

A Min Tjoa is with Institute of Software Technology and Interactive Systems, Vienna University of Technology, Vienna, Austria (e-mail: amin@ifs.tuwien.ac.at).

II. APPLICATION FIELDS

With rapid changes taking place in the field of health care, decision support systems play an increasingly important role. Health care institutions are deploying data warehouse applications as decision support tools for strategic decision making [3]. Caused by the huge data volumes (for example results of clinical trials in the epidemiological domain), data warehousing and data mining solutions are required. They provide tools to acquire medical data, to extract relevant information from it, and to make this knowledge available to all persons involved with healthcare.

Evidence-based medicine decision support systems can vary in their scope. The simplest systems are fed by data concerning diseases and best practice guidelines to support care delivery. More sophisticated systems include in addition various clinical internal and external data sources to support further decision making in the area of business management, staff management etc.

Relevant data sources for the clinical decision support systems for evidence-based medicine purposes are:

- Evidence-based guidelines (in form of rules)
- Clinical data (patient data, pharmaceutical data, medical treatments, length of stay)
- Administrative data (staff skills, overtime, nursing care hours, staff sick leave)
- Financial data (treatment costs, drug costs, staff salaries, accounting, cost-effectiveness studies)
- Organisational data (room occupation, facilities, equipment)

The most relevant application fields for data warehousing in the area of evidence-based medicine are utilized to support:

1. The generation process of the evidence-based guidelines
2. The clinicians at the point of care delivery, by making evidence-based rules available
3. The controlling of clinical treatment pathways
4. The administrative and management tasks, by providing evidence-based knowledge as well as diverse organisational and financial data.

The proposed medical data warehouse contains data from all organisational departments as well as evidence-based guidelines. This data is prepared and offered to be queried and analysed in any desired way. Clinical management is often interested in finding out, which treatments and medications led to more rapid and more economic patient convalescence. Data mining and OLAP analytical functions support business decision makers in creating the most effective business strategies that satisfy both patients' expectations and financial potential. If administrative data is available in the data warehouse, it can be combined with evidence-based medicine recommendations in order to give advice about the right number of skilled staff needed for particular medical treatments. This information can further be used for work and treatment scheduling and so to support medical decision makers in the area of human resources.

In the following three case studies will represent the benefit of the application of data warehouse and data mining in the area of evidence based medicine.

III. CASE STUDIES

A. Case study 1: Generation of evidence-based guidelines

Evidence-based medicine relies on published clinical evidence in form of books, magazines, journals, healthcare protocols, clinical trials and best practice guidelines [4]. Since it is a complex task for clinicians to gather all the necessary knowledge about given diseases, the practice of evidence-based medicine would not be imaginable without IT support. The way that data warehousing and data mining support the creation of the evidence-based rules is by providing a platform and the tools for knowledge discovery and pattern recognition. Here, large amounts of data is analysed to confirm known or discover unknown trends and correlations in the data. The discovery process in medical research benefits enormously from data mining facilities.

Fig. 1 represents the generation process of treatment rules based on the clinical evidence. Data originating from different medical sources is extracted, transformed and prepared for its loading into the existing data warehouse structure. The data warehouse contains diverse patient, clinical and pharmaceutical data. Supported by a data warehouse, medical knowledge workers are able to analyse and mine a vast quantity of available data.

Data mining techniques enable us to identify trends and to recognise best practices for the different disease treatments. Knowledge workers are aiming at finding new associations and rules which are hidden in the data so far. After the rule (in form of laboratory tests, recommended drugs, therapies or medical treatments) has been created, it needs to be examined and approved by the higher authority. In case of rejection, the proposed rule is sent back to the data warehouse for further development. If the rule is approved, it is added to the database holding evidence based guidelines. Once added to the evidence-based guidelines, rules undergo systematic reviewing process.

Schuerenberg [6] describes the governance process to develop and review clinical guidelines, as it is practiced at the University Health Network (Toronto), as follows:

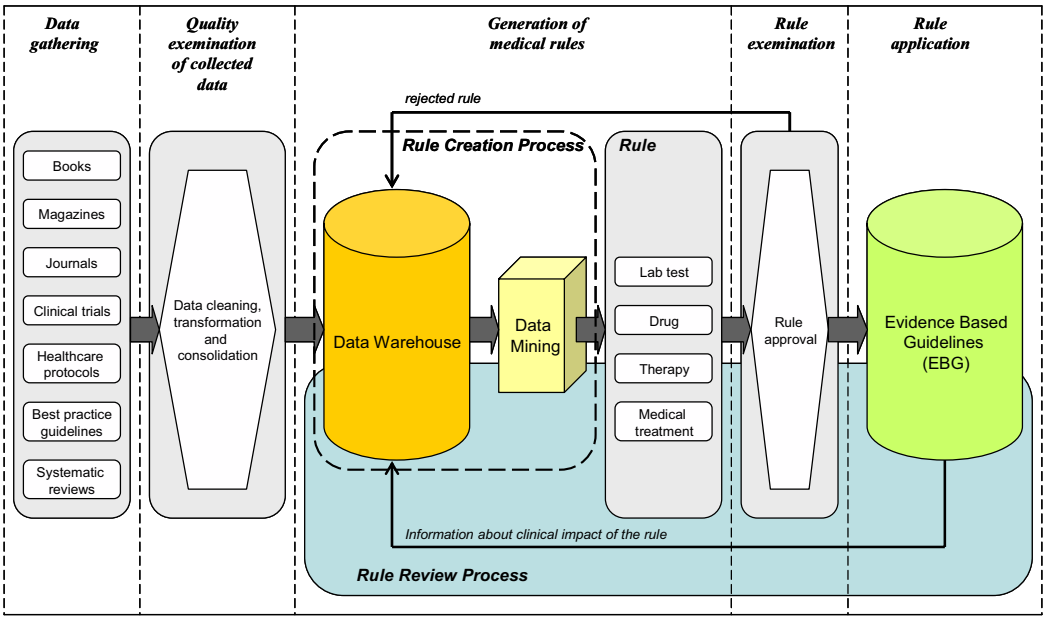

Fig. 1 Generation of evidence-based guidelines

The IT clinical advisory committee starts the process by recommending which lab tests should be ordered for which patient conditions, and why. Its recommendations are sent to an enterprise wide medical advisory committee that has final approval on which evidence-based rules and alerts are programmed into the decision support system. University Health's IT department then enters the new rule and its supporting evidence into the application. The medical advisory committee measures the clinical impact of each new automated rule six months after it's programmed. Rules often are modified during the review process and then reviewed in another six months. Development of evidence-based guidelines using data warehouse and data mining goes far beyond just revising the literature. These guidelines are the unification of the best evidence with clinical expertise and patient values, reported in sufficient detail to allow clinicians to make judgements about the validity of their recommendations and made to improve patient care.

B. Case study 2: Controlling of clinical treatment pathways

Health care organisations are searching for methods to rationalise their processes and to improve health care and ultimately also to reduce costs. As stated by the Health Informatics Research (HIR) group [11], tradition business process modelling tools and business process execution tools such as workflow management systems lack support for the complex, multi-organisation, dynamic and large scale patient treatment processes that exist within the healthcare system. New research programs are launched to develop methodologies, tools and techniques that can be applied to the more complex clinical pathway process. Clinical Pathways are structured, multidisplinary plans of care which are designed to support the implementation of clinical guidelines and protocols. They are built to support the overall clinical management, the clinical and non-clinical resource allocations, and last but not least the clinical audit and also financial management [12]. Open investigation areas include the development of new clinical pathway transformation frameworks, pathway modelling techniques and new clinical pathway process management systems.

In the area of clinical pathways, the data warehouse facilitating evidence based medicine is used for controlling of the clinical processes, from patient's admission to his (her) release. For frequent occurring diseases (like diabetes mellitus, pneumonia, hernia etc.), the whole treatment process, from diagnosis to therapy, is checked against evidence-based rules. Analysing all relevant data stored in the data warehouse, we can find out, if prescribed levels (operation, recovery duration) were reached on time. In case of significant delay, decision support system alerts the responsible clinician. The discrepancy between evidence-based rule and actual treatment can be caused by unnecessary modifications of the therapy, by the management problems but also by the incorrect rule. Evidence-based guidelines are reviewed on a regular basis (one – two times a year). Often occurring deviations of a rule are analysed and in case of legitimate causes, the rule can be reformulated. This consequently contributes to the rule refinement process.

Data mining techniques can be deployed for discovering the patterns of clinical pathways. Based on the patient record data, administrative data, clinical log data and evidence based rules, mining process is applied. With their usage, we can detect the structure of clinical paths and the sequence among activities, which human beings could hardly find. The development of clinical pathways is knowledge intensive and it requires the cooperation among knowledge workers, clinicians, nurses and clinical management. Data miners aim to combine the

experience originating from evidence-based guidelines with the concrete clinical data and identify (time) dependency patterns. Lin et al. [13] state that by obtaining the time dependency patterns, the paths for new patients can be predicted, when he (she) is admitted into a hospital, and, in turn, the health care procedure is more effective and efficient.

The goal of the use of data warehouse and data mining to support evidence-based medicine in the domain of clinical pathways is to improve the coordination of care between different clinical departments and so to improve the quality of care and to reduce the length of patient's hospitalisation.

C. Case study 3: Use of evidence-based rules at the point of care

Due to the massive economical impact of the health system, great changes in medical treatments are notable. Apart of humanitarian and healing nature of medicine, this industry is becoming more and more business like. Various serious medical studies show, that the even patients with harmless health complaints, are often exposed to unnecessary but expensive therapies. In case of serious diseases, clinicians often recommend new and costly medical treatments instead of traditional therapies. Since they are assured to get a quick recovery, patients are usually willing to try newest therapy approaches.

It is obvious, that this does not mean that new methods are not efficient. It solely means, that tests and clinical studies are needed, in order to prove their effectiveness. Evidence-based medicine offers a collection of proven best practise guidelines to recommend drugs and medical treatments. With the help of a data warehouse, clinicians can easily navigate through the knowledge database and find appropriate therapy for the particular disease.

In this case study, we show how a data warehouse is used to determine the best suitable treatment for a particular patient, using evidence-based guidelines.

The described healthcare institution is a clinic, in which the patients are treated ambulatory as well as hospitalised. Clinical business management is striving to deliver best medical treatment to the patients by applying the most effective therapies. In order to cope with future technology challenges and newest medical achievements, as well as to reduce administrative and treatment costs, this clinic is operating a data warehouse based on evidence-based guidelines, to support decision making at the point of care.

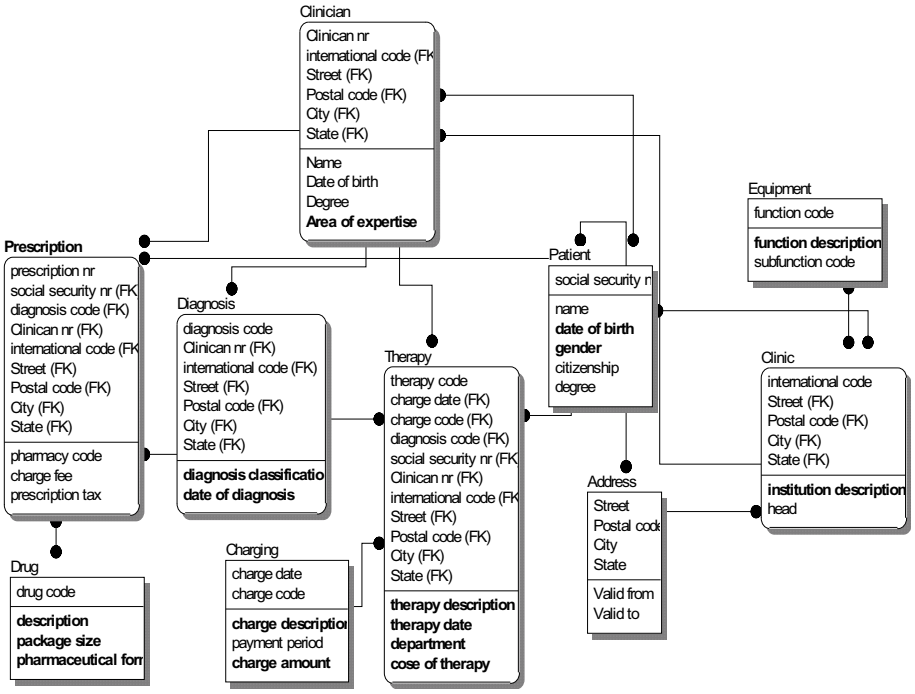

Fig. 2 Clinical data model

Fig. 2 represents an extract of a clinical data model, holding personal data about the *patient.* A patient is uniquely identified by social security number and possesses additional attributes: name, date of birth, gender etc. Patient is characterised by the relations to another entities, which may be seen as the dimensions of the particular patient. Patient may have one or more drug *prescriptions*. The set of possible drugs is given in a *drug*-entity, giving information about drug characteristics, like description, size or pharmaceutical form.

Each patient may have one or more *diagnoses*. These are specified by diagnoses codes and classifications. Since one patient can have many diagnoses, one diagnose is only valid in a certain period of time. For each patient and each diagnose made, there is a responsible clinician stored in the *clinician* entity. For each diagnose, there is one or more *therapies* assigned. Apart of therapy descriptive attributes, *charging* information is available as well. Information about the medical institutions (*clinic*), where the patients are treated and clinical facilities (*equipment*) are also available in this model.

Exceeding the data represented in this model, the data warehouse is containing the data about the staff skills and staff availability and the data about the provided facilities. An additional data source is evidence-based guidelines.

We consider the use of a data warehouse at the point of care, which means, that the clinician is querying the data warehouse while examining the patient, so the answers he (she) is expecting must come quickly and be presented clearly. The clinician is interested in finding the best fitting treatment for the patient. He (she) analyses patient's historical health data, stored in the data warehouse (entities *diagnosis* and *therapy*), as well as the *drugs*, which were prescribed so far. Patient's clinical dossier, accompanied with patient's age and lifestyle habits (originating from entity *patient*), is than combined with evidence-based guidelines. Supported by the data warehouse functionalities (aggregation, slice and dice etc.) as well as by OLAP and data mining techniques, those treatments, which proved to be the most efficient in similar cases, are detected. Since the data warehouse integrates clinic wide data from different departments, the clinician can immediately check, if the proposed treatments can be carried out (if necessary facilities (entity *equipment*) or skilled stuff is available). Having all this information, the clinician decides than, which treatment will be prescribed to the patient.

Fig. 3 illustrates the use of data warehouse facilitating evidence-based medicine, at the point of care. Firstly, clinician defines a clinical question based on the disease patient is suffering from. He uses an OLAP tool in order to query the data warehouse. Standard, predefined reports as well as ad-hoc queries can be used. After that, selected tables are joined inside the data warehouse on the fly. For example, advised medical treatments and drugs, coming from evidence-based guidelines, are matched with patient health history data, existing clinical equipment and available qualified staff, residing in the warehouse. Finally, relying on all the available data, the best fitting rule is chosen and presented to the clinician by the OLAP tool in an illustrative manner.

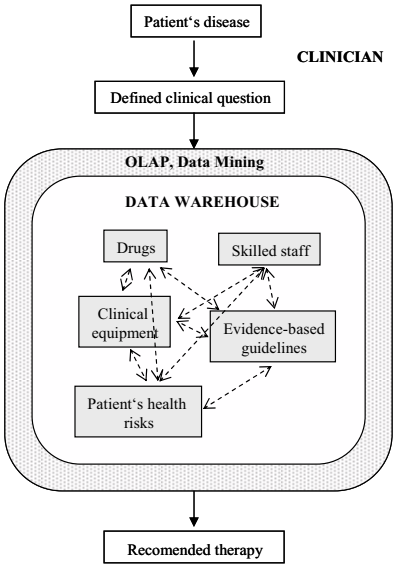

Fig. 3 Data warehouse at the point of care

IV. RELATED WORK

Abidi et al. describe in [4] an Integrated Clinical Evidence System designed to augment the typical literature–based clinical evidence with additional technology-mediated clinical evidence. They propose a technology-enriched strategy to exploit advance computer technologies – knowledge management, data mining, case based reasoning strategies and internet technology – within traditional evidence based medicine systems to derive all-encompassing clinical evidence derived from heterogeneous clinical evidence modalities.

The four steps in incorporating the best available research evidence in decision making is the subject of the research project in [9]. The authors formulate following steps: asking answerable questions; accessing the best information; appraising the information for validity and relevance; and applying the information to patient care. Further, they state that applying evidence-based medicine to individual patients requires drawing up a balance sheet of benefits and harms based on research and individual patient data. Wu et al. state in [8] that growing evidence indicates that the integration of clinical decision support into the computer-based patient record can decrease medical errors, enhance patient safety, decrease unwanted practice variation and improve patient outcomes. Clinical Pathways are the subject of research of Roeder et al., at the DRG Research Group at the Universitätsklinikum Münster [7]. They investigated 8 different international DRG-systems on the basis of data from cardiac surgery and concluded that the Australian AR-DRG-system excellently matches levels of complexity. Thus it provides a good basis for the German R-DRG-system, which will serve for the reimbursement of all in-patient cases, according to the German Ministry of Health.

Stolba et al. [10] propose a federated data warehouse approach for the use in the area of evidence-based medicine,

in order to achieve better data security. Depersonalisation and pseudonymisation are used to ensure data privacy for sensible patient data.

Data mining in the area of clinical pathways is the subject of research performed by Lin et al. [13]. They proposed a data mining technique to discover the time dependency pattern of clinical pathways for curing brain stroke. The aim of their research is to discover patterns of process execution sequences and to identify the dependent relation between activities in a majority of cases.

V. CONCLUSION

In this work we have shown the role of data warehousing and data mining technique for the use of evidence-based medicine. The consequence of not applying evidence-based medicine is the time loss in a patient treatment process. But, even when applying it, it needs to be presented to decision makers in a proper and useful manner. Only external, evidence-based knowledge is not enough for efficient treatment of individual patients. This knowledge always needs to be adjusted to the patient's health condition and preferences.

Development of evidence-based guidelines, support of the clinicians at the point of care and controlling of clinical pathways are undertakings, which can hardly be fulfilled without IT-support. Because of immense data volumes and extremely complex knowledge discovery procedures, we consider data warehouse with its OLAP and data mining tools to be a very suitable solution for accomplishment of this task.

REFERENCES

[1] Sackett D.L., Rosenberg W.M., Gray J.A., Haynes R.B., Richardson W.S., "Evidence-Based Medicine: what it is and what it isn't "[editorial]. BMJ. 196; 312 (7023) 71-72 [www.pubmed.com]W.-K. Chen, *Linear Networks and Systems* (Book style). Belmont, CA: Wadsworth, 1993, pp. 123–135.

[2] Inmon W.H., "*Building the Data Warehouse*", Second Edition, J.Wiley and Sons, New York, 1996.

[3] Shams K., Farishta M. ; "Data Warehousing: Towards knowledge management", Top Health Inf Manage. 2001 Feb, 21 (3): 24-32, [PubMed – indexed by MEDLINE]

[4] Abidi S.S.R., Abidi S.R., "A Case of Supplementing Evidence Base Medicine with Inductive Clinical Knowledge: Towards Technology Enriched Integrated Clinical Evidence System", 14th IEEE Symposium on Computer-Based Medical Systems, (CBMS'2001), 26-27 July 2001, Bethesda (USA)

[5] Frawley W., Piatetsky-Shapiro G. and Matheus C., "Knowledge Discovery in Databases: An Overview", AI Magazine, Fall 1992, pgs 213-228.

[6] Schuerenberg B. K., "Clearing the Hurdles to Decision Support", Health Data Management, May 2003.

[7] Roeden N. et al., "Clinical Pathways", Medizincontrolling/DRG Research Group, Universitätsklinikum Münster, http://drg.unimuenster.de/de/behandlungspfade/cpathways_reisebericht.html

[8] Wu R., Peters W., Morgan M.W., "The next generation of clinical decision support: linking evidence to best practice", J Healthc Inf Manag 2002 Fall;16(4):50-5.

[9] Craig J.C., Irwing L.M., Stockler M.R., "Evidence-based medicine: useful tools for decision making", The Medical Journal of Australia [http://www.mja.com.au], MJA 2001; 174:248-253M. Young, The Techincal Writers Handbook. Mill Valley, CA: University Science, 1989.

[10] Stolba N., Banek M. and Tjoa A M., "The Security Issue of Federated Data Warehouse in the Area of Evidence-Based Medicine", ARES 2006, The First International Conference on Availability, Reliability and Security, April 20.-22. 2006, Vienna, submitted for publication

[11] Health Informatics Research (HIR) group, School of Computing and IT, University of Western Sydney, (http://www.cit.uws.edu.au/hir/)

[12] Open Clinical, "Knowledge Management for medical care", http://www.openclinical.org/clinicalpathways.html

[13] Lin F., Chou S., Pan S., Chen Y., "Mining Time Dependency Patterns in Clinical Pathways", 33rd Hawaii International Conference on System Sciences, Jan 4-7 2000 Page(s):8 pp. vol.1, IEEE CNF