HEALTHCARE KNOWLEDGE COLLECTION FOR CLINICAL DECISION SUPPORT SYSTEMS

Lillian V. Ortiz-Fournier, Universidad del Turabo, marianavanessa@yahoo.com
Mysore Ramaswamy, Southern University, mysore@acm.org

ABSTRACT

This article presents several examples of how clinical decision support systems (CDSS) connected to knowledge bases help physicians and improves patient’s care. Results of a research in progress are presented, where the development of a data warehouse aims to get accurate assessments of a highly impairing mental condition. We argue that a data warehouse connected to a CDSS in the office of pediatricians will be beneficial for: physicians, patients, families, and researchers. A data warehouse was designed after careful analysis of information requirements found in literature, specific characteristics of the condition, and information available from health professionals. The theoretical framework includes the integration of a top-down with bottom up approaches to evaluate information packages and to create logical schemas. Implementation of a data warehouse is done and validation reports are designed applying a set of rules to dummy data. This system might improve children’s assessments and contributes to clinical research.

Keywords: Knowledge Acquisition Methods, Data Quality, Data Warehouses, Delphi Method, Decision Support Systems

INTRODUCTION

The Medical or Health Informatics (HI) field is mainly concerned with the management and use of information in health and biomedicine. Like the rest of Health Information Technology (HIT), is a heterogeneous field with a diversity of backgrounds, skills, and knowledge, which is probably beneficial for the complex task of working at the intersection of health care and technology [13, 14]. HI is a dynamic field of medical information, technology, and practical application, all combined in an attempt to improve health outcomes, lower healthcare costs, and educate healthcare providers and patients [24]. In addition, its core goal is to directly improve clinical care through the use of information technology. To do so, HI research examines the design of IT applications to address the practicalities of health care delivery, with a focus on clinical users (e.g., physicians, nurses, pharmacists). What is important about building this huge body of knowledge is that when developed properly information systems may improve patient health care by giving physicians better informed alternatives to backup their decisions and protect hospital health care professionals from malpractice demands.

The need to build information systems that could acquire, store, and communicate clinical data sparked the emergence of many of the great academic groups in health informatics. However, according to Blois [2] it is not the building of artifacts or the deployment of information technology in clinical settings per se that made health informatics fitting for scientific inquiry but the elucidation of the underlying medical knowledge required to build clinical information systems that demanded a theoretical foundation which is the subject of active scholarly investigation. This very essence of creating a knowledge-based system requires the construction of a model of human problem solving and the representation of human knowledge in a computational form [19]. Furthermore, the ability of professionals to diagnose mental disorders is based upon years of training, research, and experience that enable them to differentiate between possible disorders and in turn prescribe appropriate treatments. Then, like any other business, mental health care professionals and researchers need to collect, and analyze a vast amount of medical information for diagnosis and treatment purposes.

In the next sections we present the following topics: (a) why collecting data is relevant in clinical environment; (b) we then define and describe the data warehouse architecture; the proposed methodology for collecting requirements for the design of a data warehouse to keep relevant data from the Pediatricians’ Office for the assessment of Attention Deficit/Hyperactivity Disorder (ADHD); (c) preliminary results, and conclusions; and (d) further considerations, and several contributions of the data warehouse are presented.

Decision Making in the Clinical Environment

The principal purpose of computerized decision support systems in clinical practice is the support of clinical judgment providing the structures for continuing care. Computer based applications are being developed to help clinicians integrate a...
patient's preferences (values) with scientific evidence, the patient's history, and local constraints. Decision aids differ from information aids mainly in that they contain explicit components to help users clarify their values: the patient's personal values and the utility or importance of the risks and benefits of each alternative are elicited [7].

Decision support systems (DSS) are defined as information systems focused on supporting and improving managerial decision making. Arnott and Pervan [1] suggest that in a real sense, DSS is a philosophy of information systems development and use, and not a technology or homogenous field. There are a number of fundamentally different approaches to DSS and each has had a period of popularity in both research and practice.

Fox & Thomson [8] argue that the physicians’ decisions are based on the interaction between clinical goals, possible solutions and patients’ data leading to care plans and actions which at the same time influences patients’ outcomes, clinical goals and possible solutions. These interactions make the decision making process in a clinical setting a cyclic one. Purcell [20] describes clinical decision support as the integration of clinical knowledge and patient-related information, presented at appropriate times, to enhance patient care. There are several ways in which data and decision support systems are used by physicians and other health care professionals in the clinical environment. Clinical decision support systems (CDSS), as described in next paragraphs, are used for diagnosis, prognosis, pain management, patient guidance, reminders and drug alerts.

Im & Chee [16] evaluated a decision support computer program (DSCP) that can be used as an assessment support tool for cancer pain experience. Due to the great level of diversity in the United States the decision tools developed to deal with pain managements must include a knowledge base on the gender and ethnic differences in cancer pain experiences and a computer program to support decision making about cancer pain management. The development of the DSCP was based on fuzzy logic and modern control theory, and has decision rules, algorithms, and logics for classification of patients’ pain according to gender and ethnicity.

Dugas et. al. [5] developed a web-based, high-granular research database for comprehensive documentation of all relevant variables to evaluate new surgical techniques. For such a highly complicated and difficult surgery as liver neoplasm, there is enough information retrieved in huge databases about the tasks of type of surgery as well as relevant medical parameters and postoperative course. These authors describe the database as having a high level of granularity which is required for this complex research topic, and a decision support component that is a server side web application accessed using a standard web browser. The tool they developed is capable of extracting and visualizing information with a level of aggregation that enables specific insights, tailored to a specific patient, which are not available by simple database queries.

Gago et al. [9] describes the INTCare system which is an intelligent information system designed to work in Hospital Intensive Care Units designed to support the physicians’ decisions by means of prognostic data mining models. In particular, it assists the physicians’ decision making by: (i) detecting action demanding conditions by continuously scanning automatically acquired data and applying the relevant model to predict next day failure of six systems (liver, respiratory, cardiovascular, coagulation, central nervous and renal); (ii) maintaining an up-to-date in-hospital death probability value used in end-of-life decision making and (iii) evaluating scenarios for the evolution of the condition of the patient, allowing physicians to compare the consequences of different medical procedures.

Ramnarayan et. al. [21] assesses the impact of a novel diagnostic reminder system (ISABEL) on the quality of clinical decisions made by various grades of clinicians during acute assessment. Diagnostic errors impose a significant burden on modern healthcare: they account for a large proportion of medical adverse events in general, and form the second leading cause for malpractice suits against hospitals. The main hypotheses underlying the development of ISABEL were that the provision of diagnostic reminders generated following a brief data entry session in free text would promote user uptake, and lead to improvement in the quality of diagnostic decision making in acute medical settings.

Rochon et al. [23] states that sometimes, simple prescribing decisions can initiate unforeseen outcomes. A key advantage of a clinical physician order entry is that decision support can be provided when drug therapies are ordered, even remotely. They describe a clinical physician order decision support system as an automated interface that links the newly ordered drug therapy to clinical, drug and laboratory information in a way that guides prescription decisions in the direction of increased
safety, and ultimately reduces the medication errors that lead to adverse events.

The expert’s reasoning is characterized by efficient and quickly cognitive shortcuts (named “clinical reasoning skills”) triggered by a few elements. Razzouk et. al. [22] describes the development of a clinical decision support system for the diagnosis of schizophrenia spectrum disorders (SADDESQ) that includes four stages such as: knowledge acquisition, knowledge organization, the development of a computer assisted model, and the evaluation of the system’s performance. SADDESQ is a decision support system developed to help students and novice psychiatrists to understand all the necessary steps in the diagnostic decision-making process.

The implementation of effective clinical decision support is a challenging task involving interactions between technologies and organizations, and there are no easy solutions to guarantee success or to avoid failure in this complex process. Many factors influence reductions in errors or improvements in health, so measuring the effectiveness of decision support systems in improving these endpoints is difficult [20].

**Retrieving Relevant Health Related Data**

As stated by Hsieh & Lin [15] a data warehouse is a computer system capable of storing, retrieving and managing large amounts of data gathered from diverse sources. It must be dynamic, flexible and extensible in order to meet the changing requirements of users and the business as a whole. Then, a data warehouse is one way to transform data from what was stored in a transaction database system to a format that is designed to support the analyses that an organization wants [11]. The data in the data warehouse is: separate, available, integrated, time stamped, subject oriented, nonvolatile, and accessible. Therefore, before planning the implementation of a data warehouse, it is necessary to build the business case for the project which, for instance, must state such things as the types of data to be included in the data warehouse, the kinds of decisions that will be made with the aid of that data, the way those decisions are made without the data warehouse, and how decisions will benefit the organization [18]. Also, businesses can be organized along sets of hierarchical categories but medicine is unique because numerous aspects of care intertwine and overlap [6]. To manage this complexity, we need to group data into episodes of care that may be composed of multiple encounters with various providers. Clinical users of the data warehouse will need to view data that yields information about the entire patient care press.

Wyderka [27] studied an example of a health care organization in which the measurements used for clinical analysis provided the most challenges to the database design. The measurements, called “facts” in the data warehousing language, are those numbers of interest to the analyst or manager. For a business, these numbers are usually dollar amounts and for the medical outcomes data warehouse many of the facts are not meaningful when added together. There are other approaches to the development of a data warehouse for health care such as: dimensional model with an incremental development used in the CATCH project [1]; the use of large legacy snapshots with data verified using simple monitors in the WHIPS project [10]; and the Stanley Medical Research Institute online genomics database (SMRIDB) which is an online data warehouse and analytical system designed to aid researchers in understanding the biological associations both between and within the brain disorders of schizophrenia, bipolar, and major depression described by Higgs et. al. [12], among others.

**Focus of the Study**

Due to a high prevalence among children, researchers and physicians agree that an early assessment of ADHD is needed. William et. al. [26] argue that the most frequent reason of pediatricians for not making behavioral health diagnoses is related to the uncertainty of the diagnosis. Therefore, it is important to extend the availability of resources to every physician so that the best possible assessment possible. The focus of this investigation is to design a the logical schema of a data warehouse that will keep relevant data for assessment of Attention Deficit with Hyperactivity Disorder (ADHD) with or without comorbid disorders in children when visiting the pediatricians’ office. This data warehouse can be connected to a clinical decision support system. The data warehouse discussed in this paper was designed using the general architecture model for implementation of such a technology. It is aimed to receive information in different formats that is cleansed and uploaded to the data warehouse using the schema modeled. The systems requirements used to model the schema are collected using a methodology based on Bonifatti and colleagues [3] design model.
RESEARCH METHODOLOGY

Data in a warehouse can be accessed without tying up the information sources (e.g., holding locks or slowing down processing). Getting data from the warehouse does not result in costs that might be associated with accessing data at the information sources, and warehouse data is available even when the original information source(s) are inaccessible [10]. Bonifatti, et al. [3] includes research approaches of different authors for physical and logical design of a data warehouse with a focus on identifying and building data marts. They propose a design model based on the following three steps: (i) a top-down analysis consisting of collecting requirements through interviews with business analysts and/or managers about the company goals and needs; (ii) a bottom-up analysis devoted to examining the conceptual model of the operational databases, finding out candidate star schemas for the data warehouse using a graph analysis technique and the entity-relationship representation of the conceptual schema of the operational databases; and (iii) an integration process that matches each ideal star schema determined by the first step with all the candidate star schemas produced by the second step to choose the candidates of the ideal schema.

The methodology described by Bonifatti and colleagues [3] integrates two approaches to design data marts, a top-down analysis to collect the user requirements from a general business perspective, and a bottom-up analysis that aims to examine the existing operational data structures. We adapted their model to design our data warehouse. To these approaches we added a consensus technique. The Delphi technique was used to collect the data requirements from experts and intended users through the use of questionnaires delivered in two rounds. This technique is by definition a group process that includes the interaction between the researcher and a group of experts on a specific topic through the use of interviews or questionnaires [28]. The Delphi technique was selected because it has the following characteristics: anonymity of participants; iteration and controlled feedback, which allows the participants to know the opinions of the others and refine their own; and statistical aggregation of group response, that allows for quantitative and interpretation of data [25]. Our main objective is to investigate to which extent it is possible to design and implement a data warehouse environment comprised by the integration of ADHD AAP guidelines, the items of a validated assessment scale(s) and the data of the patients’ health record. The questionnaires consist of all the data items found in literature to be relevant in the assessment of ADHD and take into consideration all the health dimensions of these patients when visiting the pediatrician’s office. The data collected through these instruments might also be used to develop ideal rules for a clinical decision support system aimed to help pediatricians in the future.

A group of experts consisting of psychologists and pediatricians were selected using the snowball technique. Seven psychologists and five pediatricians accepted to participate in the study. The questionnaires for both psychologists and pediatricians contained the same data items. Psychologists were selected as experts if they had an above average knowledge of ADHD and a vast experience treating children with the condition. They evaluated the items in terms of their need for an accurate assessment of ADHD. Pediatricians on the other hand evaluated the items from the operational point of view rating the data items as being collected or not, routinely in a patient visit. On each round each expert had 10 days to answer the questionnaire. Figure 1 presents the steps carried out in two rounds of the Delphi Technique on each of the two approaches, top-down and bottom-up analysis.

The data requirements were presented to experts in tables that represent Information Packages that they had to evaluate and rate in a five point Likert scale from ‘Essential’ through ‘Not-Essential’. They also classified the items as ‘Time dependent’, ‘ADHD indicator’, and ‘Fact’ or ‘Attribute.’ In addition, demographic data and questions about behavior scales were included. After the experts finished the first round the researcher collected the questionnaires personally and had a brief meeting with them. Experts had the opportunity to discuss the questionnaire in general and describe their experience rating the data and answering the questions. The answers of the questionnaires are analyzed individually as well as in the aggregated form as they are summarized and delivered to the experts for the corresponding review (second round). The aggregated data was analyzed using descriptive statistics. The final answers are used to develop an entity-relationship (E-R) diagram to identify potential facts and dimensions. This E-R diagram is used to design the most effective schema for the data warehouse, and for modeling a set of rules for the decision aid.
Figure 1. Methodology for collecting systems requirements using the Delphi Technique

To test the validity of the logical schema an actual non-functional prototype of the data warehouse is implemented. This prototype is used to describe the ability of the system to use data from a data warehouse and rules based on expert consensus aid to provide an ADHD assessment that resembles the diagnosis given by a trained specialist. In order to comply with HIPAA no actual data from patients is collected. On its place dummy data from non-existent patients was obtained from selected experts. Using their experience and knowledge a set of dummy patients with certain characteristics was created to test how the data is uploaded to the data warehouse. A simulation was prepared applying to dummy data a set of rules that a CDSS might evaluate to help a pediatrician with the ADHD assessment. Four reports were designed which simulate the electronic patient record. The reports include data items that can be found in a patients’ record when visiting the pediatrician.
PRELIMINARY RESULTS

The questionnaires included data items found in literature relating to ADHD assessment. Their content validity index was .81 and was determined by six professionals who evaluated every item in three categories: pertinence, suitability and wording. After validation, formal invitations were given to professionals considered experts on ADHD, and pediatricians. Twenty eight professionals were invited to participate in the study but only seven psychologists and five pediatricians accepted to participate. Table 1 shows the demographics of the twelve experts who completed the questionnaires on the two rounds.

Table 1. Demographics

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Psychologists</th>
<th>Pediatricians</th>
</tr>
</thead>
<tbody>
<tr>
<td>Gender:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>Female</td>
<td>6</td>
<td>3</td>
</tr>
<tr>
<td>Age:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>25-34</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>35-44</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>45-54</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>55-64</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>Years of work:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>2-10</td>
<td>6</td>
<td>2</td>
</tr>
<tr>
<td>21+</td>
<td>1</td>
<td>3</td>
</tr>
<tr>
<td>Office Location:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Metropolitan</td>
<td>5</td>
<td>3</td>
</tr>
<tr>
<td>North</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>Center-Oriental</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>South</td>
<td>1</td>
<td></td>
</tr>
<tr>
<td>ADHD knowledge:</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Basic</td>
<td>7</td>
<td>4</td>
</tr>
<tr>
<td>Advanced</td>
<td>1</td>
<td></td>
</tr>
</tbody>
</table>

Phases 1 and 2 were carried out simultaneously since all experts, psychologists and pediatricians, were completing the first and second round at the same time. Answers collected from questionnaires in the first round were filled into a template to obtain summarized answers to each one of the data items presented as well as descriptive statistic measures using the demographic data collected. Then the questionnaires were delivered again to experts for the second round. For this round the experts reviewed their answers, comparing them with all summarized answers to give them the opportunity to refine or change their views in the light of the knowledge from the other experts. From meetings with experts we found that all of them concluded that the questionnaire was really long because it covered a vast review of literature on ADHD. However, despite the amount of data items and information that can be collected through the questionnaire most of them considered that it was relatively easy to understand and complete. The purpose of the research can be extrapolated through the questionnaire giving more validity to the results obtained from them. Experts impressions’ so far had been positive and contributes with tacit knowledge from intended users that will be very useful at the implementation phase. From few of the meetings we can cite:

“The questionnaire is very complete, really pertinent to the condition, easy to follow and to file.” (Pediatrician 1)

“The instrument has a vast amount of information for ADHD assessment. The goal here must be to build collaboration bridges between physicians and mental health professionals. At the end of the day the most important thing is the health of the patient that includes a healthy balance between mind and body.” (Psychologist 1)

Given that the instruments covered so much information, few experts expressed tiredness when completing them, and were also a little confused when classifying the data items as fact or attribute, and time dependent. In the second round, this was clarified a little more when they saw the summaries of the answers provided to them for the review.

“The questionnaire was too long, when doing research you need to be careful because you might include the analysis of too much information that is difficult to handle and this can add more confusion and distress to experts and researchers.” (Psychologist 6)

After analyzing the answers of both psychologists and pediatricians we decided to keep only the data items classified as essential, relevant and necessary. Then two schemas were developed. The psychologists’ schema, contained less items on these classifications than the pediatrician’s schema. The pediatrician’s classified all the items as essential, relevant, or necessary, respectively. The selected schema used to design the data warehouse include all the items from the psychologists’ schema and a few that were left out by psychologists but that are necessary in the patients’ record as stated by pediatricians. The result is a data warehouse that is easy to manage and its performance is not affected. An actual implementation of a non-functional data
warehouse has started and will be used to test the validity of the logical schema in light of the accuracy of an assessment derived from applying ADHD guidelines to dummy data entered into the data warehouse.

CONCLUSIONS
The purpose of this article is to present examples of how technology in the form of clinical decision support systems connected to knowledge bases can be developed and implemented in the clinical environment. Each tool has a unique task but all of them are designed with the purpose of helping physicians to improve their treatments and helping patients to get a better health care. Knowledge and theory developed within health informatics may also have a broader, cross-disciplinary role, and in the future others may explore the possibilities for interdisciplinary exchange of such research experiences to other disciplines, including information systems [4]. Therefore, it has been recognized how the introduction of technology to clinical practice can contribute to quality of care, patients better informed and clinicians more confident in their diagnoses and treatments.

The main reason to design a data warehouse system to collect data relevant to ADHD is the high prevalence of this condition among children of many countries. In addition, ADHD is not an illness that would grow out in the adolescence or adulthood [17]. Despite the existence of guidelines designed for pediatricians to assess ADHD in children, a vast majority of the ADHD assessments and diagnosis are done by psychologists, based on teachers and parents complains. The research presented on this paper includes a methodology that is somewhat different to the conventional approaches used to design and implement a data warehouse. The difference strives mainly in the fact that there is no schema developed to integrate mental health information that is collected from the patients’ record at a pediatrician’s office. For this study the data was collected through the use of validated questionnaires completed in two rounds by psychologists that are experts on ADHD and pediatricians. Experts’ opinions were also useful to validate the instruments and the methodology.

FUTURE CONSIDERATIONS AND CONTRIBUTIONS
Based on preliminary results future research can be developed to test the use of this methodology to collect requirements from other professionals that come into contact with children having ADHD indicators. In addition, this methodology can be tested with other conditions known to have high prevalence among children such as: ‘Asperger syndrome’, ‘Tourrette’s syndrome’, ‘Autism’, ‘Generalized anxiety disorder’, among others. The logical schema serves as the basis to add later other professionals that come into contact with children having ADHD indicators such as: neurologists, pediatric specialists, teachers, social workers, psychiatrists, among others.

This data warehouse can be connected to a decision support system to help pediatricians, patients, families, and researchers. Further steps include the design and implementation of a CDSS using the rules derived from the analysis of the data structure in the data warehouse, and the ADHD assessment guidelines.

The data warehouse described in this paper is aimed to improve health care and quality of life of children with and without ADHD in the following ways:
1. It is clear and free of personal bias to improve the assessment and suggestions that can be trusted and understood.
2. Help to detect at an early age the symptoms associated to ADHD; especially in the presence of comorbid disorders.
3. Provide relevant and reliable information providing the necessary confidence to pediatricians to do a referral if necessary, or to discard ADHD indicators.
4. Improve the quality of data available to parents.
5. Pediatricians will be able to extend his/her assessment to include information about different areas of the child’s life that can be associated to ADHD with or without comorbid disorders.
6. Help to relieve the pressure and the guilt from children that do not comply with adults’ expectations.
7. Give pediatricians, parents and children the possibility to design a joint treatment plan that is suitable and beneficial.

Furthermore, data collected in the proposed data warehouse will contribute to research of ADHD and its comorbid disorders in the following areas:
1. The information available from the data warehouse will help to implement better treatments, best practices, and public health policies to attend the necessities of children diagnosed with ADHD.
2. The model to collect the requirements will contribute to determine the type of information that is recorded at pediatricians’ office, its
relevance to the assessment and treatment, and identification of information needed for the assessment but not recorded by the doctors.

3. The system will provide a media for acquiring a vast amount of relevant information about children’s physical health, behaviors and attitudes, quality of life at home and school, and parents’ skills.

4. The data warehouse will serve as a repository of information that can be analyzed statistically to provide useful insights such as: condition assessment statistics, possible predictors, treatment effectiveness, referral necessity, among others.

5. This model set the basis for future inclusion of information recorded by other pediatric specialists; mental health care providers such as, psychologists, neuropsychologists, clinical psychologists, pediatric psychiatrists; teachers, and social workers.

REFERENCES


