Medical Informatics Europe '99
P. Kokol et al. (Eds.)
IOS Press, 1999

Data Warehousing as a Tool for Quality Management in Oncology

S. Hölzer, A.G. Tafazzoli, U. Altmann, W. Wächter, J. Dudeck
Institute of Medical Informatics, Justus-Liebig-University Gießen, Heinrich-Buff-Ring 44, D-35392 Gießen, Germany

Abstract: At present, physicians are constrained by their limited skills to integrate and understand the growing amount of electronic medical information. To handle, extract, integrate, analyse and take advantage of the gathered information regarding the quality of patient care, the concept of a data warehouse seems to be especially interesting in medicine. Medical data warehousing allows the physicians to take advantage of all the operational data they have been collecting over the years. Our purpose is to build a data warehouse in order to use all available information about cancer patients. We think that with the sensible use of this tool, there are economic benefits for the Society and an improvement of quality of medical care for patients.

Keywords: Data Warehouse, Analytical Processing, Quality Management, Oncology, Tumour Documentation, Medical Informatics

1. Introduction

The intelligent management of information is essential to the efficient operation of any large enterprise or institution. Particularly in the health care domain with its dynamically changing environment this management has become important to provide up-to-date patient care. The increasing use of computers by the medical profession as well as in the administrative area has led to large amounts of structured and unstructured, clinical and non-clinical data. To co-ordinate the flow of information and to distribute it adequately hospital information systems have been developed. On the other hand, the health professional has to perform more and better clinical work with stagnant resources. Our purpose is to build a data warehouse in order to use all available information about cancer patients. This tool seems to be especially beneficial to evaluate actual patient care and optimise the provided quality of care.

2. Material and Methods

In this context we intend to outline the general concept of a data warehouse, demonstrate its possible application in health care and medical research, and point out current problems. Subsequently, we explain the idea and give practical examples dealing with typical data of oncological patients as a very specialised category of medical data.
Medical knowledge is increasing very fast. Large amounts of clinical and non-clinical (e.g. administrative) data are routinely collected into computer systems and are available for analysis under certain conditions. This growing infrastructure can improve the

management of health care and the understanding of disease if adequate tools to extract relevant information are developed. At present, physicians are constrained by their limited skills to integrate and understand growing amount of electronic medical information. They lack an essential enzyme to digest this material. The volume of data is irrelevant if it isn't organised so that it adds value. Value to an organisation means turning data into actionable information. Value to a medical institution means turning data into information which can improve quality of patient care, understanding of disease, and economic use of resources. Data warehouses, a term of the business and financial world, manage large amounts of historical and reference data for decision making and reporting [1,2]. They consist of three major components. Tools to extract, transform and load operational and external data sources; a database to store and manage the data; and end-user tools to reference and analyse the data.

Information about patients, their medical history and disease specific items are collected at different levels, e.g. hospitals, general practitioners, clinical registries, epidemiological registries, pharmacies, health insurance companies, public health centres and so on. To gather this spread information is a big challenge. Only data available in a structured form can be useful for central analytical processing. Gaining access to previously unavailable information sources, building documentation facilities and infrastructure (e.g. hospital information systems, distributed electronic medical records) are first steps in creating a data warehouse. This means that the success of a data warehouse depends to a great extend on the quality of the underlying databases [3]. This informational system (data warehouse) is different from the operational systems such as e.g. hospital information systems. Our intention is to optimise the collection of data of oncological patients from various sources (operational systems) into our data warehouse. The management of tumour patients almost always includes a multicentric and interdisciplinary approach. This way, information from these patients is collected in the hospital by physicians of surgery, radiology, radiotherapy, pathology and by the administration (Admission, Discharge, Transfer). Other, external information comes from the above mentioned institutions. Tumour centres gather these information sources and transfer them to our central database. But the integration of medical data is still a big problem. There is no application consistency in encoding, key structure and naming convention. Some data exist in many places with different names; some data are labelled the same in different places but are still the same and some data are in the same places with the same name but reflect a different measurement of data ....

As data pass from the operational environment to the data warehouse environment, they are integrated.

3. Results and Discussion

Whereas at the institutional level (hospital) communication standards such as HL 7 or EDIFACT are well established, a data exchange format called BDT (abbreviation for the German word: Behandlungsdatenträger) exists at the level of ambulatory care in Germany. In order to facilitate in part this process of integration, an extension of this special data exchange format for oncological data was designed. A basic data set for cancer patients is also available in Germany (Basisdokumentation für Tumorkranke). On the basis of this Domain Information Model the above mentioned extension of the data exchange format was defined [4].

Thus we are able to easily collect the information deriving from operational databases of hospital departments, tumour registries and so on by using the standardised data exchange format. But the integration of other sources rests problematic. In a pilot project we are

collecting data for patients with a special tumour (thyroid carcinoma) in cooperation with the Commission on Cancer of the American College of Surgeons in the United States: Information from both National Cancer Databases (NCDB in the US and Germany) plus data from cancer centres involved in this project are gathered. This way, we are combining information of cross-sectional studies and specially designed prospective studies. A data warehouse is time variant because it is associated with specific time periods [5]. The time horizon of our data is about 8 years. Data warehouse data is snapshots taken as of some moment in time. We load data in different cycles (monthly or annually). In medicine there is a strong need to have the relevant information topical and up-to-date. As more applications will come on-line in the future more frequent refresh cycles are possible. Following that, the concept of the data warehouse offers different ideas and techniques to manage and analyse data that make it especially useful for physicians.

As mentioned above, data warehouses are separated logically and physically from operational systems and optimised for the retrieval of information. It has the advantage to being accessible by health professionals on their own terms, regardless of their understanding of underlying data structures. The benefit consists of viewing the integrated data, of running large queries without impacting operational systems and of analysing "what if" scenarios. The analysis of time trends based on historical information is essential in medical research. For example, a gynecologist can recognise trends in breast cancer staging or shifts in the use of different treatments (surgery, radiotherapy, adjuvant chemotherapy ...). One can get clinical, epidemiological, social or even economic data of a geographic region, a period, a tumour stage selected for patients with certain risk factors or other conditions. This process is known as drilling down. Drill downs can continue until the user gets hold of the necessary information.

Data can also be sliced into user-defined views. This makes it possible to view and to work with it in multiple dimensions. This multidimensional database usually contains a star model which represents data as an array where each perspective is a subject around which analysis is performed. The data warehouse is oriented to the major subject areas that are in our case the patient or the quality of care for the patient. Getting the answer to typical questions from raw data of an oncological patient often requires such viewing from various perspectives. For example, a physician of nuclear medicine, treating patients with thyroid carcinoma, wants to know which groups of his patients (depending on histological type, stage, age, sex) get a real benefit (survival time, time of remission, post-treatment side effects and complications ...) of a selected therapy (applied radio-iodine activity, number of cycles ...). To overcome the constraints of small numbers of patients in the generated subgroups even for rare neoplasms, information sources of long term follow-up are combined with data of regional and time limited cross-sectional studies.

Additional options of a data warehouse include ranking, sorting, exception filtering and visualisation. At the information access layer we are using standard applications such as Microsoft Excel and SPSS statistical packages for analysing and visualisation. One has to take into account that the medical knowledge of associations between the parameters is complex and different sorts of bias increase the probability of misinterpretations. Open Database Connectivity (ODBC) and Standard Query Language (ORACLE and Microsoft ACCESS databases) are important components of the data access layer. But the technical details are not contents of this paper.

In the future we intend to organise our central tumour database to become a data warehouse with more and more up-to-date (online) and clinical relevant information. The development of features that will provide analysis, query, and reporting for the health professional to access useful information quickly and easily seems to be the most difficult and sophisticated step. Integrating more and more data of better quality from distributed

sources will improve the validity of such a tool. Thus, data warehousing could become essential for the quality management in medicine [6-8].

4. Conclusion

Medical data warehousing allows the physicians to take advantage of all the operational data they have been collecting over the years. These data are a valuable corporate asset, yet seem seldom to be used strategically. The great value of analysing large quantity of medical information is to illuminate hidden problems and to point to their solutions. The interesting feature of data warehousing is its ability to provide the answers to questions not yet asked, by finding trends, patterns or anomalies in data. This feature could become very valuable in handling a growing amount of medical data. With its sensible use there are economic benefits for the Society and an improvement of quality of medical care for patients.

Acknowledgements

This project is supported by the German Federal Ministry of Health.

References

[1] Neil R. Technology Tutorial 1 - Data warehouse. *Information Week* 1996; No. 571
[2] Neil R. Technology Tutorial 2 - Data warehouse. *Information Week* 1996; No. 572
[3] Holthuis J, Muksch H, Reiser M. Das Data Warehouse Konzept - ein Überblick. *Wirtschaftsinformatik* 1996; 38: 421-433.
[4] Altmann, U, Wächter, W, Müller, A, Funken, O, Sembritzki, J, Lichtner, F, Gehlen, E, Dudeck, J. Datenaustausch in der Onkologie mittels BDT. *Medizinische Informatik, Biometrie und Epidemiologie.* Greiser,E, Wischnewsky, M (Editors) 1998, 83: 247-250.
[5] Mundy J. From the ground up: Building a data warehouse. *Sybase Server Magazine* Fall 1995
[6] Hölzer S, Dudeck J. Patient Care Evaluation Studies: Applying this concept of quality management in oncology. *Medical Informatics Europe* 1997; 834-838.
[7] Hölzer S, Dudeck J. Patient Care Evaluation Studien als Werkzeug der Qualitätssicherung in der Onkologie - Ist eine Übertragung auf Deutsche Verhältnisse möglich? *Biomedical Journal* 1998; 51:8-11.
[8] Hölzer, S, Dudeck, J., and Clive, R. Realising a Co-ordinated Approach to Quality Management in Oncology. Cesnik, B., McCray, A. T., and Scherrer, J. R. *MedInfo ,Proceedings of the Ninth World Congress on Medical Informatics* 1998, 890-890.