Data and Knowledge for Medical Decision Support
B. Blobel et al. (Eds.)
IOS Press, 2013

doi:10.3233/978-1-61499-240-0-130

Project I-COP – Architecture of Software Tool for Decision Support in Oncology

Milan BLAHA[1], Dalibor JANČA, Petr KLIKA, Jan MUŽÍK and Ladislav DUŠEK
Institute of Biostatistics and Analyses, Masaryk University, Brno, Czech Republic

Abstract: This article briefly describes the development of the I-COP tool, which is designed to promote education and decision making of clinical oncologists. It is based on real data from medical facilities, which are processed, stored in database, analyzed and finally displayed in an interactive software application. Used data sources are shortly described in individual sections together with the functionality of developed tools. The final goal of this project is to provide support for work and education within each involved partner center. Clinical oncologists are therefore supposed to be the authors and users at the same time.

Keywords: I-COP, data processing, data mining, oncology, data analyses, medical education, National Cancer Registry, insurance companies reports

Introduction

Each medical facility has a very complex structure that cannot operate without the support of a number of larger and smaller systems. These systems are primarily focused on supporting diagnostic and therapeutic processes in a hospital, but the support of ongoing administrative processes is equally important. The main source of information is undoubtedly the Hospital Information System (HIS). In all these systems, and this is particularly true for HIS, a large amount of useful data is being collected on the treatment of patients. Unfortunately, it is usually not used for re-evaluation of such treatment.

Our project, in cooperation with 11 cancer care centers (CCC) in the Czech Republic, strives to overcome this fact and use administrative data to support clinical activities of oncologists. This involves processing of hospital administrative data, addition of structured clinical data and creation of a related database. Such database is then used as the basic building element for the described application, which generates analytical reports and allows medical doctors to see the previously hidden parts of the data, display them in a clinically relevant way and use it for further evaluation, analysis and education.

[1]Corresponding Author. Milan Blaha, Ph.D., Masaryk University, Institute of Biostatistics and Analyses, Kotlarska 2, 611 37 Brno, Czech Republic; Email: blaham@iba.muni.cz; Phone: +420549495047; URL: www.iba.muni.cz

1. Methods

1.1. Hospital administrative data

HISs contain a lot of valuable information, though their direct and uniform usage for analysis is problematic. Various hospitals run different HISs, which usually contain data in differently structured form. Moreover, HIS data are not always readily available at a reasonable cost to the operator. Therefore, we decided to use insurance companies' reports (ICR) as a source of administrative data [1]. These reports are available at each hospital, always date a few years back and are completely independent of a specific HIS.

Technically, ICR are plain text files with a defined structure given by the methodology of VZP ČR. This structure is, however, variable in time, which has to be taken into account during their processing.

ICR can be used for each individual patient to find particular information about medical procedures and administered medicaments during hospital and outpatient care.

In I-COP, ICR are anonymized, birth IDs are replaced by secure salted hash codes and data are processed in accordance with database dictionaries maintaining history of their structure and facilitating extraction of their contents. These data are stored in the first layer of an I-COP database, which is fully described in the following paragraphs. In principle, ICR describe complete process of care for patient in a facility, although from an operational point of view and while neglecting some specific details.

1.2. National Cancer Registry data

The National Cancer Registry (NCR) is a structured database [2], which, by law, records every newly diagnosed neoplasm in the country since 1976. This epidemiological database contains basic clinical parameters in a structured form, such as diagnosis and stage of disease, that affect the prognosis and treatment of patients and associated costs. The register contains records of initial treatment, but only in general and not entirely thorough form. For example, we will not find data on the administration of targeted anticancer therapy in the registry. The greatest limitation however is the fact that the register covers only the primary diagnosis and treatment process and the data on subsequent treatment of patient is totally inadequate.

Each medical facility may request the registry administrator, the Institute of Health Information and Statistics of the Czech Republic (ÚZIS), to export records of relevant cancer patients from the registry. Export is implemented in the form of CSV text files, where one line carries information of one diagnosed neoplasm. Individual parameters are defined on an attached data interface.

1.3. The combined database

The proposed solution is to combine both data sources into a single database, where the administrative reports of individual cancer patient in particular medical facility meet with the patient's record from the National Cancer Registry. This combined database is trying to eliminate the disadvantages of both the aforementioned data sources and use them in a complementary system [3, 4, 5].

In practice, it is implemented in MySQL. The first database layer holds both data sources in nearly raw form. The second layer is generated by connecting these sources

together, namely by transformation leading to the identification of coherent therapeutic segments of oncological patient treatment. In this connection, it is possible to identify and describe not only the primary care for cancer patients, but it is also possible to focus on the analysis of relapse or progression of cancer in its later stages.

The combined database is the basis for definition of further outputs, which are briefly described in the following sections. Generally, the registry tells us who we treat and why, while the ICR show us how we treat. I-COP database provides a separate data source that allows us to perform specific exports and realize advanced analytical outputs.

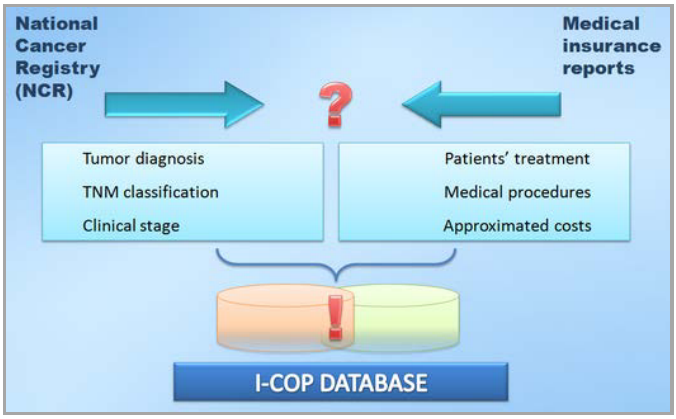

Figure 1. Schema of combined I-COP database.

2. Results

2.1. I-COP Reporting

An analytical system is being built above the combined database, which presents its individual outputs through a web application. The first part of the application, I-COP Reporting, is designed as a standardized set of predefined reports that allows us to view basic epidemiological trends and compare information about patients of particular medical facility with the reference values of the Czech Republic or its region [6]. The selected group of patients may be further restricted by a set of dynamic filters, which allows us to precisely define the analyzed data set and to display the desired report.

The entire application is being implemented using modern technologies, making it as user-friendly as possible. All its modules are loaded dynamically and reflect state of the combined database as-is, in real time. Most of the application controls provide users with detailed information in the form of extensive tooltips. Some reports are also interactive allowing the users to work with their content on the fly.

Figure 2. I-COP Reporting filters and graphs.

2.2. I-COP Browser

The Browser tool is integrated in an I-COP web application and is intended for advanced oncologists, to support their education and scientific work. The logic of the tool is aligned with I-COP Reporting and includes a set of dynamic filters allowing selection of specific patients. The range of the filters is wider though, because patients can be selected on the basis of applied treatment as well. For example, it is possible to select only patients with a given diagnosis who underwent specific medical procedure and pharmacotherapy, all at the same time.

The output of this tool is in the form of pivot table, which contains patients matching conditions of the selected filters. Each cell represents the number of patients meeting certain criteria. Overall, this makes it a powerful tool while searching for a group of patients for a specific study. Individual table cells are interactive – clicking on each cell displays a list of patients meeting the criteria. It is also possible to go through to the level of an individual patient, specific medical procedure or therapeutic modality.

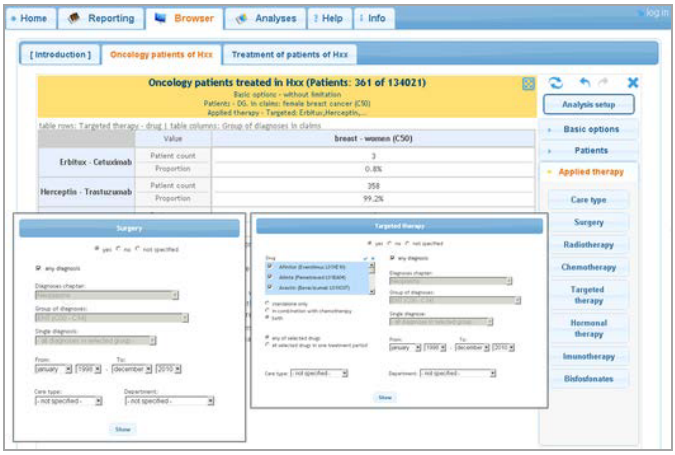

Figure 3. I-COP Browser filters and pivot table.

Treatment of each patient can be visually displayed on an interactive chart that maps the entire process and shows its individual components, including quantification of costs based on the reported price points and medicaments used. This way, I-COP Browser allows selection of a user defined category of cancer patients and provides a way to follow through, up to the details of an individual patient.

3. Discussion

Hospital administrative data were used for mapping the process of care for cancer patients. The description of this process in conjunction with the clinical data from the NCR brings a whole new data source and a new insight into the data on the treatment of cancer patients, now available for clinical oncologists. The combined data are processed in a joint database and tools are being developed that will allow the data to be displayed and analyzed. In collaboration with clinical oncologists, an analytical tool was designed, which has reached an advanced stage of development. This tool will be very useful for evaluation of treatment in a medical facility, allowing oncologists and hospital management to take appropriate action.

The possibility of obtaining, processing and using oncological data for the support of decision making of related medical doctors was tested and verified on a model case of administrative and parametric data. The proposed analytic application is operational and is further expandable. In its later development, we would like to focus on sharing the outputs from the application between participating centers and further enrich its contents by integration of other data sources, such as laboratories or other complements.

A very important aspect of the I-COP project is the versatility of the technological solution that can be applied in other fields of medicine, particularly in places where there is a structured parametric data source available. For these reasons, our further thoughts aim to use the proposed application in the field of cardiology.

Acknowledgement

The works has been supported by the Ministry of Education, program OP VK, CZ.1.07/2.4.00/31.0020.

References

[1] VZP ČR. Datové rozhraní individuálních dokladů. 2012, http://www.vzp.cz/uploads/document/dr2-v6221.pdf
[2] ÚZIS ČR. Národní onkologický registr. 2010-2012, http://www.uzis.cz/registry-nzis/nor
[3] Earle C, Nattinger A, Potosky A, Lang K, Mallick R, Berger M, Warren J. Identifying Cancer Relapse Using SEER-Medicare Data. Medical Care 2002 Aug; 40(8, Suppl):IV75-IV81.
[4] Zheng H, Yucel R, Ayanian J, Zaslavsky A. Profiling Providers on Use of Adjuvant Chemotherapy by Combining Cancer Registry and Medical Record Data. Medical Care 2006 Jan; 44(1): 1-7.
[5] Nathan H, Pawlik T. Limitations of Claims and Registry Data in Surgical Oncology Research. Annals of Surgical Oncology 2007 Nov; 15(2):415-423, DOI: 10.1245/s10434-007-9658-3
[6] Dušek L, et al. Czech Cancer Care in Numbers 2008-2009. Praha: Grada Publishing; 2010, ISBN 978-80-247-3244-2.