Web portal for information on cancer epidemiology in the Czech Republic

Dusek L.\textsuperscript{1,2}, Muzik J.\textsuperscript{1}, Kubasek M.\textsuperscript{3}, Koptikova J.\textsuperscript{1}, Brabec P.\textsuperscript{1}, Zaloudik J.\textsuperscript{4}, Vyzula R.\textsuperscript{4}

Abstract

The aim of this article is to inform about newly issued automated web portal focused on population risk analyses related to cancer epidemiology. Portal was built up on very representative database of National Cancer Register (Czech Republic, Ministry of Health Care, standardized collection of cancer data from 1997 - 2005) that provides fully representative long-term trends. Nowadays the database consists of more than 1 200 000 cases stratified according to main risk factors and diagnostic descriptors including TNM classification of tumours. The automated system of on-line analyses offers unique access to cancer registries, demographic and environmental databases. Portal is principally developed as tool increasing the information potential of risk assessment studies. Portal is available at http://www.cba.muni.cz/svod/.

1. Introduction

Cancer epidemiology can be regarded as one of the most important and most frequently analysed topic in the field of human risk assessment. It is not only due to remarkable public concern about growing population risk, cancer incidence and mortality is evident and clearly reachable endpoint for risk assessment studies. We can enter this problem from the viewpoint of risk factors as agents initiating carcinogenesis, but epidemiological parameters can retrospectively indicate hazardous impact on population of a large scale. The bioindication from epidemiological data of course requires sufficient data sources. It means representative long-term profiles of incidence and mortality and a very good awareness of most important risk factors.

Analysing epidemiological trends we must be able to distinguish between statistically significant trends and random fluctuations. For risk assessors it would be most important to recognize environmentally related cases that can be attributed to external factors like pollution of air, drinking water and/or food. And again, the influence of these factors must be filtered on the substantial background of the other “natural” risk factors like age structure of the population, genetic factors or frequently omitted life style. Apart from these complicated circumstances we can concentrate our attention to some cohorts of oncological patients that might probably indicate external impact – if they increase in incidence and non randomly in some regional or temporal scales. It means namely occurrence of less advanced disease stages in age groups that are commonly out of main risk (for example breast carcinoma cases in pre-menopausal women younger than 35 – 40 years). In that context, we can take information benefit mainly from diagnostic groups that

\begin{enumerate}
\item Centre of Biostatistics and Analyses, Faculty of Medicine and Faculty of Science Masaryk University in Brno, Czech Republic
\item Research Centre for Environmental Chemistry and Ecotoxicology, Faculty of Science, Masaryk University in Brno, Czech Republic
\item Department of Information Technologies, Faculty of Informatics, Masaryk University in Brno, Czech Republic
\item Masaryk Memorial Cancer Institute, Brno, Czech Republic
\end{enumerate}
might be related to some external influence (colorectal carcinoma, kidney and lung carcinoma, breast cancer, malignant melanoma, …).

All these analyses require easily available large data sets that are itself very expensive and typically not directly available – epidemiological cancer registries. In addition to it, we must aggregate cancer data with demographic data in order to attain for example age-specific profiles of incidence. That is why there is growing interest of many professional groups (health care managers, environmental experts, risk assessors) in accessibility of these data. In our experience however, the demand for data cannot be easily fulfilled by blind databases of primary population data. The analyses are very time-consuming and finally, the outputs might be ambiguous and not safe to be communicated with public. Therefore, we developed professional web portal that offers automatically generated and verified epidemiological analyses on cancer incidence and mortality in Czech Republic. This presentation is aimed to introduce potential users to the technology, architecture and functioning of the portal.

Web portal is accessible through the following address: http://www.cba.muni.cz/svod/. It is developed in Czech and English version.

2. Aims of the portal and accessible data sources

The portal is principally aimed to provide several automated analyses that are in detail specified in chapter 4. The analyses were developed over several data sources that are automatically aggregated according to user’s choice (cancer epidemiology data, demographic data, data on risk status of Czech population…). All analytical functions result in fully colored graphical and table protocols that are opened for further editing and modification. The portal functions are targeted primarily for health care managers and risk assessors working in the field of human and ecological risk assessment. Some types of outputs (namely incidence profiles) are prepared in a very safe way and are widely accessible to general public.

Portal was built up on very representative database of National Cancer Register (Czech Republic, Ministry of Health Care, standardized collection of cancer data from 1997) that provides fully representative long-term trends. Nowadays the database consists of more than 1 200 000 cases stratified according to main risk factors and diagnostic descriptors including TNM classification of tumours.

3. Technological background, security and access rights

The whole web portal is placed on two servers. The first is database server with operation system Linux and data management system MySQL and the second is Linux based web server with PHP scripting module. Both servers are connected via reserved network interface with private IP addresses, so database server is separated and secured from whole Internet.

Special functions of web portal are accessible in secure area (access requires appropriate login and password) and all communication is based on Web Services standards (XML, SOAP, WSDL) and realized via SSL secure protocol. User interface was designed and realized by a combination of interrelated techniques like HTML templates, JavaScript and CSS. The main scripting language used for building dynamic web pages is PHP.

Management of web portal has two levels: Web-Master with full access to all functions of the portal (main administrator) and administrators with access to specific parts of the portal (management of news, software SVOD section etc.)
4. Automated analyses

Automated analyses form the core functions of the system. The analyses are accessible both directly and in mutual combination according to strategy of the user. Through these functions, you can easily analyze epidemiological trends examined over three decades, stratify and filter cohorts of patients and extract population risk in absolute or age-specific values. Some of analyses offer benchmarking with respect to clinical status of the disease. Major epidemiological trends are available in comparison with international data (GLOBOCAN 2002: Cancer Incidence, Mortality and Prevalence Worldwide).

4.1 Automated analytical tool

These tools enable to analyze data about specified diagnostic group:

**Incidence and mortality:** time trends of incidence, mortality and mortality/incidence ratio.
Available parameters are absolute numbers, crude rate (number of cases per 100000 people in population) and age standardized ratio (ASR - European or World standard)

**Time trends:** changes of incidence and mortality in time.
Available parameters are growth index related to selected year and between-years changes. Both parameters could be viewed as absolute numbers or as relative proportions.

**Regional overview:** comparison of incidence and mortality in regions of the Czech Republic.
Available parameters are crude rate and age standardized ratio (ASR - European or World standard), output could be bar-plot or map.

**Age analyses:** age structure of population of patients with selected diagnose
Available parameters are absolute numbers of cases in age categories, % of cases in age categories and age specific rate (number of cases in age category per 100000 people in population cohort of the same age).

**Clinical stages:** time trends of proportion patients in specific clinical stages.
Available parameters are absolute numbers, % and crude rate of patients in specific clinical stages; available outputs are time trend bar-plot, time trend line-plot or pie chart of selected time period.

**International data:** comparison of incidence and mortality in the Czech Republic (CZ) with other countries. All these analyses are based on data obtained from IARC database GLOBOCAN 2002.

**Comparative standards:** time trend of incidence or mortality in selected region in comparison with situation in whole Czech Republic.
Available parameters are crude rate and age standardized ratio (ASR - European or World standard).

**Typology of patients:** comprehensive overview of group of patient with specific diagnose

Examples of automated analytical tools settings and outputs are on figures 1 - 3.
Diagnose selection and consequent output window with tools for setting of the analysis, selection of specific group of patients and viewing tabular outputs and reports.

Left window: example of analysis setting – selection of requested epidemiologic parameter (incidence and/or mortality) and setting of appropriate units (absolute numbers, crude rate and age standardized ratio - European or World standard); Right window: example of tools for selection of target group of patients – available parameters are sex, age group, region, time period, clinical stage, TNM classification and other parameters related to the status of the patient.
4.2 Reporting system

Results of all automated analytical tools are available in graphical and tabular form, but these outputs are viewed in separate windows. Thus, there was developed reporting system, which enables to make comprehensive outputs of every analysis (fig. 3). These reports contain graphical and tabular outputs with detailed description of setting of the analysis and are downloadable in pdf format.

Ministry of Health Care Czech Republic supports development of Web portal SVOD. Research and developmental teams are granted by research project MZO 00209805 solved in Masaryk Memorial Cancer Institute, Brno. Risk assessment analyses are supported by research project INCHEMBIOL, Ministry of Education Czech Republic project no. 0021622412.

Bibliography
