Analysis of Population Cancer Risk Factors in National Information System SVOD

Mužík J.¹, Dušek L.¹,², Pavliš P.¹, Koptíková J.¹, Žaloudík J.³, Vyzula R.³

Abstract

Human risk assessment requires analysis of multiple data sources to make correct interpretations of the results. One of these sources is population epidemiologic databases. In this presentation we introduce System for Visualization of Oncological Data (SVOD, version 6.0) built up over the database of National Cancer Registry of Czech Republic. Today in this system are accessible data of all malignant diagnoses (C00 - C97) from time period 1977-2002 (over 1.2 million cases). Such a database represents a great pool of information that can be used in retrospective evaluation of population risk. We will present tools and functions of the system, which make this information easily accessible for wide spectrum of users.

1. Introduction

Human risk assessment in standard terminology simply means to estimate the incidence of death and disease resulting from exposure to hazardous agents. However this definition does not fully cover all the aspects associated with risk assessment studies. In practice, many times we have to analyze trends and epidemiological data without evidence that changes are directly related to some type of environmental exposure. Instead of causal analysis of population data (that are frequently limited in extent, quality and accessibility) we often enter environmental databases that are focused on hazardous compounds and exposure and analyze the problem from this point of view. Thus, ecological risk assessment serves as scientifically credible evaluation of environmental exposure that obviously can affect also human population. Many toxic and hazardous events in environment can be used as early warning entry to human risk assessment studies.

It would be however counterproductive to argue against standard epidemiological analyses and to propose some single methodical approaches based on ecological risk assessment procedures. Our presentation is aimed to demonstrate how to solve problems with accessibility of epidemiology data and how to solve often very complicated analysis and interpretation of these data. We would like to present here population epidemiology data as quantitative base for comparing and prioritizing risks, both for retrospective and prospective risk assessment studies.

The study will introduce widely accessible expert system (SVOD, version 6.0) built up over the database of National Cancer Registry of Czech Republic, that consists of more than 1 200 000 cases and the data are guaranteed by standardized methodology of collection since 1977. Such a database represents a great pool of information that can be used in retrospective evaluation of population risk, both in absolute terms

¹ Centre of Biostatistics and Analyses, Faculty of Medicine and Faculty of Science Masaryk University in Brno, Czech Republic
² Research Centre for Environmental Chemistry and Ecotoxicology, Faculty of Science, Masaryk University in Brno, Czech Republic
³ Masaryk Memorial Cancer Institute, Brno, Czech Republic

26.08.2005, MužíkJ.doc
(incidence, mortality) and in relative, population-adjusted indices. The system is developed as analytical tool that provides 4 types of automated expert analyses
- standard epidemiology survey and trend analyses (incidence, prevalence, mortality, trend indices)
- risk factors recognition and quantification (analyses standardized with respect to structure of target population, evaluation of demographic and social risk factors)
- regionally specific analyses that transform epidemiology data into regionally or locally specified cohorts (these analyses are opened for aggregation with environmental data)
- analyses for management of health care facilities (analysis of survival, diagnostic and therapeutic algorithms, etc.)

Architecture and functions of the software is presented in chapter 4 of this article. Presentation of the expert system is available at http://www.cba.muni.cz/svod/.

2. Cancer epidemiology as a model for risk assessment studies

Human risk assessment studies often consider carcinogenicity because cancer is more dreaded by the public. Apart from its societal value, cancer epidemiology provides ideal environment for development of population model and risk assessment algorithms. Cancer incidence and mortality can be studied from the viewpoint of some inherent risk factors associated with genetic constitution of each individual, but in addition to that we must take into consideration many other influential factors like life style factors, demographic and social structure of population and of course environmental impacts. That is why epidemiological studies sometimes very resemble retrospective ecological risk assessment as it frequently starts by observing of some degradation, i.e. decline in population size of some species.

Cancer epidemiology has become one of the most important topics in internationally guaranteed information systems focused on human risk data. Recent international data about cancer epidemiology are available in these databases: CI5 I-VIII (Cancer Incidence in Five Continents, Vol. I - VIII), GLOBOCAN 2002 (Cancer Incidence, Mortality and Prevalence Worldwide) and ACCIS (Automated Childhood Cancer Information System). These databases are accessible on pages of International Agency for Research on Cancer - IARC (http://www-dep.iarc.fr/). Czech national information and expert system presented here was developed to be compatible with these international systems.

3. Data model and structure of database

Data model of expert system SVOD consists of two main parts:

a) population demographic data - absolute numbers of men and women in five-year age groups in regions and districts in years 1977-2002. These data are used in all epidemiologic calculations.

b) data of National Cancer Registry of Czech Republic
These data represents collection of all cases of malignant neoplasms detected in years 1977-2002 (completely 91 diagnoses C00-C97, D03, D05 and D06 – over 1.2 million cases). Data are separated into individual databases according to diagnose of malignant neoplasm, each record represents newly detected malignant neoplasm in population.
Individual parameters are related to these main areas:
- **basic parameters about patient** – sex, year of birth, age at diagnosis, region and district, social status, main area of employment, smoking and occurrence of malignant neoplasm in relatives
- **diagnostic parameters** – diagnosis, year of diagnosis, laterality of neoplasm, detailed TNM classification, clinical stage, topography of neoplasm, histology of tumour and grading,
- **clinical parameters** – used diagnostic methods, used therapeutic methods (surgery, radiotherapy, chemotherapy, hormonal therapy, other types of therapies)
- **parameters about patient status** – current status (alive/death), cause of death, date of death

Updates of all databases are made yearly.

## 4. Architecture and functions of the software

### 4.1 Used technology

Software SVOD is working with Windows operation systems (98/ME and above). It is written in C++ programming language, communication between software and database is done via SQL language and OLEDB interface. Data of individual diagnoses are stored in the Microsoft Access 97 database format. Software is also prepared for communication with Microsoft SQL Server database.

### 4.2 Function of the software

The main function of the SVOD software is focused on the analysis and the interpretation of the population cancer data.

**Diagnosis selection** is the first and basic step of each analysis. User can select one of 91 available diagnoses sorted into these groups:

I. Tumours of head and neck (diagnoses C00, C01, C02, C03, C04, C05, C06, C07, C08, C09)
II. Tumours of digestive organs (diagnoses C10, C11, C12, C13, C14, C15, C16, C17, C18, C19, C20, C21, C22, C23, C24, C25, C26)
III. Tumours of respiratory and intrathoracic organs (diagnoses C30, C31, C32, C33, C34, C37, C38, C39)
IV. Tumours of bone and soft tissues (diagnoses C40, C41, C45, C46, C47, C48, C49)
V. Tumours of the skin (diagnoses C43, D03, C44)
VI. Breast tumours (diagnoses C50, D05)
VII. Gynecologic tumours (diagnoses C51, C52, C53, D06, C53, D06, C54, C55, C54, C55, C56, C57, C58)
VIII. Urogenital tumours (diagnoses C60, C61, C62, C63, C64, C65, C66, C67, C68)
IX. Tumours of central nervous system and eye (diagnoses C69, C70, C71, C72)
X. Tumours of lymphoid, haematopoietic and related tissues (diagnoses C81, C82, C83, C84, C85, C88, C90, C91, C92, C93, C94, C95, C96)
XI. Tumours of endocrine glands (diagnoses C73, C74, C75)
XII. Other tumours (diagnoses C76, C77, C78, C79, C80, C97)

After diagnosis selection user can analyze data by three types of analytical tools, which are accessible by the standardized module COBRA® (Comprehensive Data Browser):
a) **Presentation** – is predefined comprehensive set of analytical outputs, which describes selected diagnosis in various points of view: time trends of incidence and mortality, gender comparison, age structure of patients and deaths caused by diagnosis, regional comparisons and clinical stages. Presentation includes comments, additional information and access to specialized analytical tools of the system (fig. 1).

![Figure 1. Example of comprehensive presentation windows in software SVOD.](image)

b) **Data browser** – this tool enables analysis of individual parameters of the database. User can perform his own analysis of selected group of patients (filtering parameters are region, period, age, sex, detailed diagnosis ICD-10, clinical stage, and other parameters related to diagnostics and treatment) and make stratified outputs. All outputs are available in tabular and graphical form (fig. 2).
c) Expert services – analytical tools fully controlled by user and focused on specific area:

**EPIDEMIOLOGY:**
- Epidemiology: incidence and mortality – time trends of incidence, mortality and mortality/incidence ratio (absolute numbers, crude rate - number of cases per 100000 people in population, age standardized ratio (ASR) - European or World standard)
- Epidemiology: changes in time - changes of incidence and mortality in time (growth index related to selected year and between-years changes - absolute numbers or percents)
- Age structure of patient population - age structure of patients or deaths caused of diagnose (absolute numbers of cases in age categories, % of cases in age categories, age specific rate - number of cases in age category per 100000 people in population cohort of the same age)
- Age specific trends in time – time trends of age-specific incidence or mortality
- Clinical stages - time trends of proportion of patients in specific clinical stages (absolute numbers, percentage or crude rate)
- Time trends of ASR – time trends of ASR with confidence intervals for regional comparisons
- Regional data: districts – tabular outputs for selected district (absolute numbers, crude rate and ASR with confidence intervals)

**COMPARATIVE ANALYSIS**
- Comparative analysis: epidemiology - time trends of incidence or mortality in selected region in comparison with situation in whole Czech Republic
- Comparative analysis: age structure – age structure of patients or deaths caused by diagnosis in selected region in comparison with situation in whole Czech Republic
- Comparative analysis: survival analysis – comparison of survival analysis outputs for selected region and Czech Republic
- Comparison of ASR in regions – comparison of ASR for all regions in selected time period

**HEALTH MANAGEMENT**
- **Diagnostics and treatment: time trends** – time trends of application of diagnostic and therapeutic methods (absolute number and percentage)
- **Diagnostics and treatment: age of patients** - application of diagnostic and therapeutic methods according to age of patients (age structure of patients with applied method(s) – absolute numbers and percentage, percentage of application of method according to age)
- **Diagnostics and treatment: combinations of methods** – tabular outputs of used combinations of diagnostic and treatment methods, enables comparison with user defined health care standards
- **Survival analysis** – tool for performing Kaplan-Meier survival analysis on selected groups of patients

All analyses can be performed on selected groups of patients (filtering parameters are region, time period, age, sex, detailed diagnosis ICD-10, TNM, clinical stage, and other parameters related to diagnostics and treatment). Outputs are available in graphical and tabular form, graphs are fully editable (fig. 3).

![Figure 3.
Example of expert services analytic window. Left window: basic setting of the analysis and selection of target group of patients; right: analytic window with graphic and tabular outputs](image)

All graphical and tabular outputs of Data browser and Expert services can be exported into other formats for use in other Windows applications (TXT and XLS for tables, BMP, JPG and EMF for graphs).

Moreover, software SVOD serves not only as an analytic tool for the assessment of population cancer data, but offers additional communication and information services:
- access to the [National web portal for cancer epidemiology in the Czech Republic](http://www.cba.muni.cz/svod/)
- [on-line internet discussion club](http://www.cba.muni.cz/svod/) – this club is assigned for SVOD users, serves as place for discussion about actual problems, enables communication with software developers and offers on-line help.
- [tutorials for analyses and user manual](http://www.cba.muni.cz/svod/).

Development of system SVOD is supported by Ministry of Health Care Czech Republic (and 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.