HEALTH STATISTICS IN INTERNATIONAL DATABASES 
AND THEIR CARTOGRAPHIC VISUALIZATION

Radim Štampach
Laboratory on Geoinformatics and Cartography, Department of Geography, Faculty of Science, Masaryk University, Brno, Czech Republic

Edvard Geryk
Department of Economy, University Hospital Brno, Brno, Czech Republic

Manuscript Received: February 29, 2012
Revised Version: July 31, 2012


ABSTRACT. For health system management, relevant data and information are necessary. The first part of the paper describes databases of the important health data providers WHO, OECD, IARC and Eurostat. It analyses whether they are in discrepancy and whether it is possible to combine data from different databases. Cartographic visualization can be used for the analysis of situations and also for the publishing of statistics and other information. Selected international providers offering health maps are evaluated in the second part of the paper. There are differences in the possibilities offered to the user and in the quality of the provided maps.

KEY WORDS: health statistics, statistical databases, health cartography, cartographic visualization

Radim Štampach, Laboratory on Geoinformatics and Cartography, Department of Geography, Faculty of Science, Masaryk University, Kotlarska 2, 61137 Brno, Czech Republic, e-mail: stampach@mail.muni.cz

1. Introduction

The transformation of the world into a ‘global village’ brings advantages and disadvantages. One of the possible risks is a pandemic outbreak of a contagious or infectious disease. Nowadays it is possible to travel across continents in just a few hours. Under these circumstances, it is very difficult to stop the spread of an epidemic. In a pandemic situation, authorities must react very quickly and decisions must be taken in a short time. Accessible and accurate data is a requirement for accurate decisions to be made.

GIS and spatial analysis play important roles in the management of a crisis such as a pandemic situation. They allow, for example, the modelling of the likely spread of a disease or the localization of emergency infrastructure. Cartography and maps can be used for the monitoring, prediction and analysis of a situation in space as well as
time. It can be documented on many atlases with medical (mostly oncological) topic – e.g. Konečný et al. (2008).

Not only pandemic outbreaks but also lifestyle diseases are threats to modern society, the global economy and general wellbeing. A well-known example is cancer. The IARC Lyon Globocan database estimates that 10.8 million new cases of cancer (excluding non-melanoma skin cancer) were diagnosed worldwide in 2002 and 12.7 million in 2008. 6.7 million patients died of cancer in 2002 and 7.6 million in 2008 (Ferlay et al. 2010, Štampach et al. 2010b). The number of cases is increasing year on year. A similar situation can also be found with regard to diabetes and other chronic lifestyle diseases.

Such phenomena place stress on the health care system. A higher number of cases requires greater expenditure on health care. State budgets, however, are not increasing sufficiently to match the increasing number of cases of cancer, diabetes and other diseases. The situation has become something of a ‘time-bomb’, and there is only one possible solution: we have to put a much greater emphasis on prevention, which means, largely, helping people to change their lifestyles. Early diagnosis of cancer increases the probability that therapy will be both successful and cheaper. The general public must be informed about this issue and the possibilities open to them (Štampach et al. 2010a).

Advertising campaigns require accurate and available data and must also be disseminated to the general public in an interesting and readable way. Maps can be ideal for this purpose. A map is a picture; therefore, it can be easily used in various media, such as newspapers, the Internet or TV. A map is also usually more interesting and visually pleasing to a user than a table with numbers.

It was stated above that reliable and accessible data are important for health system planning and that maps are useful tools in the areas of health situation analysis and crisis situation management as well as in information dissemination and health lifestyle propagation. This paper focuses on the review and evaluation of important sources of data and maps with a health theme accessible on the Internet. The first part of the paper describes accessible health databases and the second part concerns maps with a health theme. It should be mentioned that both parts consider resources offered by the same organizations. This is because most health statistics providers also offer their data in the form of maps.

2. International health data providers review

Nowadays, there are many databases containing international health data. Databases of the most important health data providers – WHO, the Regional Office of WHO for Europe (WHO/Europe), OECD, IARC and Eurostat – are described in this part. The review focuses on the most important databases administered by the abovementioned providers. These databases contain many attributes and values collected from many countries.

The following characteristics were identified for each of the reviewed databases:

- number of countries covered by the database – a higher number of countries allows better analysis of the international situation and better analysis of the situation in a selected part of world.
- the level of administrative unit that the statistics were collected for – statistics available for smaller administrative units than the state allow better analysis of the spatial distribution of phenomena.
- the number of years that the statistics have been available – a longer time period for available values allows better analysis of trends in time.

We were also interested in whether there were some discrepancies between various datasets and whether it was possible to combine values from different databases. The comments in the next section are very short because databases develop very quickly. The number of attributes and the number of years these attributes are available increase month by month; therefore, any complex review would quickly be obsolete. In the description of each dataset, we emphasize the characteristics that are special for that particular database and make it different from other sources of health statistics.
3. Existing databases and their statistics

A summary of relevant information regarding the various data sources (link, available spatial and temporal extent) is presented in Table 1.

3.1. World Health Organization  
– http://www.who.int

WHO and its regional offices collect and provide a large number of statistics in many databases. Only the most important of them are described here.

**World Health Statistics**  
– http://www.who.int/whosis/whostat/

An annually published report (WHO WHS 2011) with statistics about the health status of inhabitants of the 193 member states of WHO. The book is in the form of a .PDF file; the statistics can also be downloaded in an .XLS file.

<table>
<thead>
<tr>
<th>Name</th>
<th>Uniform Resource Locator</th>
<th>Region, Number of Countries</th>
<th>Level of Regions</th>
<th>Years</th>
</tr>
</thead>
<tbody>
<tr>
<td>WHO</td>
<td><a href="http://www.who.int">http://www.who.int</a></td>
<td>World</td>
<td></td>
<td></td>
</tr>
<tr>
<td>World Health Statistics</td>
<td><a href="http://www.who.int/whosis/whostat/">http://www.who.int/whosis/whostat/</a></td>
<td>193</td>
<td>state</td>
<td>various</td>
</tr>
<tr>
<td>Global Health Observatory</td>
<td><a href="http://www.who.int/gho/">http://www.who.int/gho/</a></td>
<td>193</td>
<td>state</td>
<td>various</td>
</tr>
<tr>
<td>WHO Global Infobase Online</td>
<td><a href="http://www.who.int/infobase">http://www.who.int/infobase</a></td>
<td>193</td>
<td>state</td>
<td>various</td>
</tr>
<tr>
<td>Global Health Atlas</td>
<td><a href="http://apps.who.int/globalatlas/data/query/default.asp">http://apps.who.int/globalatlas/data/query/default.asp</a></td>
<td>193</td>
<td>state</td>
<td>various</td>
</tr>
<tr>
<td>WHO/Europe</td>
<td><a href="http://www.euro.who.int">http://www.euro.who.int</a></td>
<td>Europe + Central Asia</td>
<td></td>
<td></td>
</tr>
<tr>
<td>European Hospital Morbidity Database</td>
<td><a href="http://data.euro.who.int/hmdb/">http://data.euro.who.int/hmdb/</a></td>
<td>27</td>
<td>state</td>
<td>1999–2010</td>
</tr>
<tr>
<td>IARC</td>
<td><a href="http://www.iarc.fr/">http://www.iarc.fr/</a></td>
<td>World</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Cancer Incidence in Five Continents</td>
<td><a href="http://ci5.iarc.fr/">http://ci5.iarc.fr/</a></td>
<td>78</td>
<td>state (more values for some states)</td>
<td>various</td>
</tr>
<tr>
<td>OECD</td>
<td><a href="http://www.oecd.org/">http://www.oecd.org/</a></td>
<td>World</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eurostat</td>
<td><a href="http://epp.eurostat.ec.europa.eu/">http://epp.eurostat.ec.europa.eu/</a></td>
<td>Europe</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Notes:
1) Number of countries is an approximate value. It varies between various characteristics in database.
2) The first and last year in the database. The time series of most states and attributes are much shorter.
Global Health Observatory
– http://www.who.int/gho/

This portal (WHO GHO 2011) provides data, materials and analysis about a very wide range of topics. Many attributes are included: mortality and the burden of diseases, children’s health, risk factors, HIV, tuberculosis, immunization, world health statistics (incidence, prevalence, health care expenditures) and many others.

WHO Global Infobase Online
– http://www.who.int/infobase

Tables, country profiles and other information about chronic diseases (alcoholism, obesity, diabetes etc.) are provided here (WHO GI 2011).

Global Health Atlas
– http://apps.who.int/globalatlas/dataQuery/

There are various statistics in the form of tables and maps. The user can choose theme, geographical area and years (WHO GHA 2011). Data can be downloaded in .XLS format.

3.2. World Health Organization – Regional Office for Europe
– http://www.euro.who.int

The European office of WHO (WHO/Europe) manages the greatest amount of data of all regional offices of WHO. Data are supplied by institutions that manage health statistics in each country. The portal of WHO/Europe provides access to a broad range of databases. Only four of them were considered in this paper.

European Health for All Database
– http://www.euro.who.int/hfadb

This is the main WHO/Europe database (WHO/Europe HFADB 2011). There is a broad range of data – demography, mortality, health care resources etc.

European Detailed Mortality Database
– http://data.euro.who.int/dmdb

This database (WHO/Europe DMDB 2011) enables comparison of the causes of death in different age groups separated by 5-year increments. The user can compare various causes of death in one selected country or compare situations in more countries for one selected cause of death.

European Hospital Morbidity Database
– http://data.euro.who.int/hmdb/

Hospital care statistics (number of patients, length of hospitalization etc.) for various diagnoses are contained here (WHO/Europe HMDB 2011). The user can search various statistics in one country or compare the situation in more countries for a selected characteristic.

European Mortality Database
– http://data.euro.who.int/hfamdb/

This database (WHO/Europe EMDB 2011d) provides data on the causes of death. In contrast to the European detailed mortality database, data for regions smaller than the state are also provided (strictly speaking NUTS 2). The problem is that data for NUTS 2 regions is provided only for 12 states including Russia and Uzbekistan; many European countries are, however, missing – e.g. Germany and Italy.

3.3. International Agency for Research on Cancer
– http://www.iarc.fr/

IARC provides many databases dedicated to cancer statistics. Two of them are described here: the cancer epidemiology databases Cancer Incidence in Five Continents and Globocan.

Cancer Incidence in Five Continents
– http://ci5.iarc.fr/

This database (Curado et al. 2007) contains data on cancer incidence (new cases occurring in members of a population in a specified time period). Incidence is presented as number of cases, crude rate per 100,000 inhabitants, and the age-standardized incidence rate standardized on the World Standard Population (ASR-W). One complication in using these data is that numbers are not collected for states but for cancer registries. It means that Spain and Britain, for example, have more values in the databases, because there is more than one registry in that country.
Globocan 2008
– http://globocan.iarc.fr/

The Globocan 2008 database (Ferlay et al. 2010) contains the global estimation of cancer incidence and mortality. The values for each kind of diagnosis are provided as the number of cases, crude rate per 100,000 inhabitants, the age-standardized incidence rate standardized on the World Standard Population (ASR-W) and the cumulative risk of dying from, or developing cancer before the age of 75 in 184 countries in 2008. Each value represents one state. Statistics for 28 different kinds of cancer diagnoses can be found in the form of tables and maps. Crude and relative rate, age-standardized rate and predictions up to the year 2030 are provided. It should be emphasized that values are only estimations. All states of the world are represented in the databases, including countries where data is rarely collected. E.g. incidence values for Afghanistan are calculated from values pertaining to Tajikistan and the Kerman province of Iran.

3.4. Organization for Economic Co-operation and Development
– http://www.oecd.org/

OECD collects a great number of data; health statistics are part of them.

OECD Health Data
– http://www.oecd.org/health/healthdata/

A large number of attributes is accessible after paying a fee (OECD HD 2011). It is updated annually. Some basic indicators (number of doctors, risk factors, health care expenditures etc.) can be downloaded for free in an .XLS file.

Health Statistics
– http://stats.oecd.org/

This online database (OECD HS 2011) contains similar health data as the free .XLS file mentioned above. But there are also statistics about financing of public health systems. The user can find economic data on health expenditure and the provider and funding of health care – data which is hard to find in other databases.

3.5. Eurostat

The Statistical Office of the European Union (Eurostat 2011) manages a wide spectrum of data, including health statistics. Types of collected data with a health theme are defined in Regulation No. 1338/2008 of the European Parliament and of the Council on Community statistics on public health and health and safety at work (EU 2008). Health data are concentrated in two sections:

Health

This section consists of two parts. Basic statistics about life expectancy, causes of death, number of patients and doctors etc. come under Public Health. Information about accidents and hazards at work etc. come under Health and Safety at Work.

Regions and cities

This section contains various data on regions smaller than the state (for NUTS 2 regions). Health statistics are represented by numbers of doctors, dentists and hospital beds. Values are provided for states of the EU and some other countries, e.g. Switzerland. Unfortunately, not all data are complete for all countries – e.g. the number of doctors is not known for NUTS 2 of Germany.

4. Discussion

It can be said that many international health datasets are accessible on the Internet, most of them containing a wide range of characteristics for many countries. We did not find discrepancies in compared datasets during our review. However, combining data from different sources can be problematic.

Values are frequently computed not for real but for theoretical ‘standard’ populations in order to allow comparison of countries with different age structures in their populations. The problem is that different databases use different standard populations – e.g. the OECD uses...
the OECD population for 1980, WHO and IARC use the World Standard Population, WHO/Europe and Eurostat use the European Standard Population (Štampach & Geryk 2011). The same attribute for the same country can have a different value in a different dataset. Some values are only estimates, e.g. statistics from the Globocan 2008 database. In short, all these abovementioned problems show that it is very important to study the attribute definition and method of data collection before using statistics from any dataset.

It is important also to mention some other facts that must be considered by the user. Definitions of collected attributes change over time; this means that, sometimes, statistical time series are not homogenous. This can complicate using values to predict situations in the future. The definitions of attributes can differ among different countries sending values to international databases. Some states do not provide all attributes – e.g. some attributes for Germany are missing in the Eurostat datasets. Data for administrative units smaller than the state are important for complex spatial analysis; however, there is a lack of such values. Only WHO/Europe and Eurostat provide them for NUTS 2 regions, but only for a few countries.

5. Health map providers review

Cartographic visualization is a possible method for presenting health statistics, including data from international health databases described in the previous part of paper. Cartography can also be used for monitoring and analyzing a situation – in space as well as in time. Health maps, however, are not only used by professional cartographers. They are also used by medical specialists and by the general public. Maps which are too complex or confusing can cause users, especially non-cartographers, to misinterpret their data. This led us to investigate various statistical health maps offered on the web.

International health map providers are often the same as the health data providers mentioned in the previous chapter, e.g. WHO, WHO/Europe, IARC and Eurostat. Indeed, most health statistics providers also offer their data in the form of maps. Most providers use interactive web map tools on their web pages but classical static maps also exist. We reviewed several maps that are accessible on the internet and investigated the quality of the provided maps, the possibilities provided to the user, and the way the maps were published. Differences between the compared providers are commented on in the next section.

6. Available maps a map applications

The following section does not offer an exhaustive list of organizations offering health maps. It includes examples which best represent the differences between providers.


The main WHO source of maps is the web page http://www.who.int/globalatlas (WHO GHA 2011). Two kinds of maps can be accessed from this web page. Static maps (Maps and Resources link) and interactive maps (Interactive Mapping link).

Maps and Resources

This is the name of an extensive archive of classical maps prepared for print, which covers various themes, regions and years. The maps can be downloaded in .PNG and JPEG format. Their graphical quality is mostly very good. An example is shown in Figure 1.

Interactive Mapping

This application is an inversion of a previous archive. The user chooses a theme, time and region and produces his own map (Fig. 2). There is a possibility to combine more than one attribute in the map. Accessible attributes can be divided into categories: contagious diseases, non-contagious diseases, health system employers and indicators of health status and health care.
6.2. World Health Organization – Regional Office for Europe
– http://www.euro.who.int

Two of the above-mentioned databases of WHO/Europe – the European Health for All Database (WHO/Europe HFADB 2011) and the European Mortality Database (WHO/Europe EMDB 2011) – can visualize their data in map form. The user has only limited possibilities to change the produced map. He can specify the classification type and change interval borders.

An example of this kind of map is shown in Figure 3. The map is not satisfactory from a cartographic perspective.

Notes:
1) Number of countries is an approximate value. It varies between various characteristics in database.
2) The first and last year in the database. The time series of most states and attributes are much shorter.
tographic point of view. The blanking out of states outside Europe can be confusing – Israel looks like an island. There is also a further problem. Figure 4 shows the same map opened in the
Mozilla Firefox or Opera web browsers. The legend is deformed and the map is shifted. This map tool works well only with Internet Explorer. In our opinion, the map tool of such an important database should not be affected by the type of browser used. It must be added that more than 25% of users used Mozilla Firefox or Opera in November 2011 (Statcounter 2011). This means that the great part of potential users would see a deformed map. This problem with web browser compatibility was not seen in other web map tools described in this paper.

6.3. International Agency for Research on Cancer  

Data from the Globocan 2008 database (Ferlay et al. 2010) described above can be visualized in map form (Fig. 5). Cartographic problems were not found. The user has only limited possibilities to change the map but he can choose the colour.

6.4. Eurostat  

The Eurostat portal (Eurostat 2011) allows visualization of its data in an interactive map. In contrast with WHO/Europe and IARC maps, the Eurostat application offers many possibilities to change map appearance. The user can choose colour, number of intervals and interval borders. The statistical distribution of values is also very useful (Fig. 6). It is the only map application from those described in this paper that offers an alternative method to the choropleth map. Figure 7 shows the use of graduated symbols. Values can be shown in the map; the user can change the number and borders of intervals.

Fig. 5. Map of estimated age standardized mortality for colorectal cancer in 2008 (Ferlay et al. 2010).
7. Discussion

The facts described above show that there are differences among data providers with regard to health data visualization. Of the described providers, only WHO offer classical static maps. Most organizations publish their statistics by means of cartographic web tools connected to values in a database. Some of these tools allow only limited changes to be made to the produced map, e.g. those used by IARC or WHO/Europe. By contrast, the tools employed by WHO and Eurostat are highly interactive – the user can influence many features of each map.

The most frequent type of cartographic visualization is the choropleth map. All the reviewed tools offer it. It is a simple but powerful and well-known cartographical method of displaying information. Cognitive studies with complex experiments (e.g. Pickle 2003) show that non-cartographers prefer this type of map. However, it is worth remembering that, from a cartographic point of view, the choropleth map is not ideal for all types of attributes. The visualization of raw values, e.g. ‘number of doctors’, in choropleth map form could cause the erroneous interpretation of a situation. This means that there is a need to offer also alternative cartographical methods for visualizing health data. However, only Eurostat offers another possible kind of visualization in its map tool – graduated symbols (Fig. 7).

The quality of the offered maps differs from provider to provider, but in general there are no significant problems. There is one exception, however: maps created by the WHO/Europe web tool are, from a cartographic point of view, unsuitable for data presentation (Fig. 3). Furthermore, this web tool works well only with the Internet Explorer web browser; maps are deformed in other web browsers.

Fig. 6. Death due to diabetes mellitus per 100,000 inhabitants in 2005 visualized by choropleth map (Eurostat 2011).
8. Conclusion

The number of sources providing data and maps with a health theme is increasing, as well as the number of attributes provided by each source. The problem lies in the incompatibility of different sources. The use of different definitions and methods of standardization does not allow the combination of data from different providers. There is a lack of data for administrative units smaller than the state (i.e. NUTS 2 or NUTS 3 regions) and too few economic indicators with respect to health systems.

Health data can be used in crisis situation management (to analyze the current situation and to predict a situation) as well as health system management (for planning health care provision in the future). However, to make data useful, they must be presented in an efficient way which is meaningful to both medical professionals and to the general public. A good way of presenting data about health status and health care is to display them in map form. International data providers also offer their data in the form of maps; however, the quality of such maps and the possibilities for the user to modify them are very different from source to source.

References


