Approaches Towards Health Information

Karl A. Froeschl, Wilfried Grossmann, wilfried.grossmann@univie.ac.at
Institute for Scientific Computing, University Vienna

Wolfgang Dorda, Georg Duftschmid, Walter Gall
Informatics Group at the Medical University Vienna

Markus Moschner
Technical University Vienna

Abstract

The paper discusses two different approaches towards health information, one is the so called statistical approach and the other one is the approach used in electronic health records. Differences between these two approaches are outlined and possible benefits from integration of these two approaches are discussed. A conceptual framework for integration of the two approaches is proposed.

Introduction

The development of eHealth systems, which plays a prominent role within the eEurope initiative of the European Union ([4]), offers a new powerful source for health information about citizens. Inside such eHealth systems the so called electronic health record (EHR) plays an important role. ISO/DTR 20514 defines the basic generic EHR as „... digitally stored health care information about an individual’s lifetime with the purpose of supporting continuity of care, education and research, and ensuring confidentiality at all times.” Due to this definition main emphasis was up to now on flexible data interchange between different non centralized or distributed EHR systems in connection with secure and confidential storage of patient records, in order to facilitate communication between health professionals about the status of a person ([2], [6], [7]). Hence, the data are characterized by a medical view onto individual patients and in this paper we will denote all kind of such data, informing about patients as EHR-data.

Besides EHR-data there is traditionally a second important data source about health, which is based on health statistics, taking a summarizing view onto health for an underlying population. Such health statistics are usually based on specific surveys or on register information from administrative data about health. In order to inform about the health status in society and to support health policy and health research the data are organized not as individual records about patients but as a set of interrelated tables about illnesses, symptoms and the health system. We will call them in this paper as S-data

Up to now the development of these two approaches is more or less separated, but in recent times a number of authors have proposed a closer connection between these two types of health information. In this short note we want to present some basic ideas for building a stronger relationship between EHR-data and S-data. In order to build such a bridge between the two sources we discuss in section 2 the main differences between the two approaches towards health information and section 3 shows some benefits of better integration between EHR-data and S-data. In section 4 we discuss a conceptual model for health data which is compatible with both approaches and outline the basic steps necessary for obtaining integrated health information from different data sources.

1 Comparison between EHR-data and S-data

A coherent common usage of EHR-data and S-data is only possible if we have a clear picture about the conceptual differences in the two different types of data. The following four topics, which are an immediate consequence of the two different approaches towards health information, are of utmost importance

a) Underlying population and representativity

Contrary to the medical view, which focuses mainly on individual patients the statistical view looks more on the population of a country or a subgroup of interest like women, persons in a specific age group, or persons with specific symptoms. Integration of the data will be only meaningful if we can identify in both data sets the coincidences and the differences in the population of interest. In general we can assume that EHR-data cover
only a subset of the entire population which is not selected according to criteria of representativeness in the statistical sense, but according to the occurrence of medical interventions. S-data fulfill usually the representativity criteria according to some criteria, but are by no means complete with respect to medical interventions.

b) Attributes and Variables

Usually the attributes in EHR-data give much more detailed information about the health status of a person with respect to clinical parameters. For example in EHR-data one could find detailed information about the blood pressure of a person, in many cases as a time series of measurement records together with information about the treatment in case of high blood pressure. In S-data such detailed information will rarely occur. In the example of blood pressure one would usually obtain from a survey only a summarized statement whether the person suffers from high blood pressure.

c) Method of data collection

EHR-data are usually obtained as the result of a medical intervention and are recorded by health professionals. In this sense they are an objective diagnosis or measurement. Contrary to such objective information S-data are usually reported by persons and give a subjective statement about the health status of the surveyed person with respect to the topics in the questionnaire. Such statements may be different from medical diagnosis, in particular in all cases where one is interested in information about wellbeing of a person, which can hardly be captured by objective criteria. This shows also that in some cases one cannot rely only on one method of data collection.

Another point in case of data collection is the different role of time. EHR-data give a rather detailed temporal picture of all health interventions in the lifetime of a person, whereas S-data obtained from a survey refer to the health status at a specific time point.

d) Data Structure

Like most statistical data the structure of S-data is rather simple and normalized and can be characterized by the well known case by variates matrix. In case of EHR-data the structure may be rather complex and contains besides records of measurements also many other data types, for example semi-structured data like written reports, or multimedia data files like x-ray images.

2 Benefits from Integration of EHR-data and S-data

Recently a number of authors have discussed benefits from using EHR-data in connection with public health issues ([1], [3], [8]). Let us mention the following topics.

a) Augmentation of S-Data by EHR-data

S-Data based on health surveys give many times information about the health status of a person using a rather general medical terminology without details. For example, persons may be asked whether they suffer on diabetes without details about type of diabetes and method of therapy. Such details may be found in EHR-data and using statistical modeling techniques it may be possible to combine the information from the survey with appropriately aggregated EHR-data in order to draw a more detailed picture about the health status of the entire population.

b) Reduction of Response burden

In some cases it may be possible to infer all the necessary information about the health status of a group of persons from EHR-data. For example in case of operations or in case of carcinomas we can expect that all information is captured in EHR-data, together with the socio-demographic information of interest. Using this information may reduce the amount of questions presented to the surveyed persons, which means reduction of the response burden.

c) Comparison between subjective and objective wellbeing

As already stated in section 2 there is a strong subjective component in all health issues. EHR-data and S-data are with respect to this subjective component rather complementary with EHR-data focusing more on the objective criteria of health and S-data more to personal well being. A comparison of the two data sources can give us insight into the discrepancies between the two definitions of health. Another topic of interest is all kinds of treatments which run besides the official health system and are not captured in EHR-data, for example treatment of spinal pains.

d) Medical research

Evaluation of the implications of the results of medical research for health policy depend many times essentially on a combination of data from clinical trials with structural data about health status of the population, i.e. data captured as EHR-data and data captured as S-data.
3 A conceptual framework for data integration

Due to the fact that EHR-data offer in general data at a personal level and S-data are at an anonymized population level, an integration scenario makes no sense at an individual level but only in aggregated form at a population level. Such integration has also the advantage that we can avoid all problems of privacy with respect to personal information.

Figure 1: Schematic integration of EHR-data and S-data

Figure 1 show in schematic way how the two data sources can be integrated. The entire population $P$ is split into two parts $P'$ and $P''$, where $P'$ are those persons for which information in EHR-data is available and $P''$ are those persons for which no EHR-data exist. The persons from which information in S-data is available is marked by the ellipse $S$, which is also split into two sets $S'$ and $S''$ corresponding to the partitioning of $P$. S-data are collected in the data store $DS_s$ and afterwards aggregated in a datamart $DM_s$. EHR-data are collected in a data store $DS_{p'}$ and afterwards aggregated in a datamart $DM_{p'}$. Even in case of same dimensions for both datamarts and same summary attributes the two cubes are not comparable due to the two sets $A$ and $B$, which represent those persons for which EHR-data exist but no S-data ($A$) and those persons which have been surveyed but no EHR-data exist ($B$).

The size of these two sets $A$ and $B$ depend on the method of data collection, the subject matter question of interest and the dimensions of the aggregation in the datamart (cf. section 2). Depending on the specifications we have to define appropriate models for making EHR-data and S-data comparable.

Due to the rather complex and different structures of EHR-data it is in general not so easy to define immediately from which EHR-data one should build the datamart $DM_{p'}$. In order to resolve this problem it seems useful to define an intermediate data structure for health data which can be understood as a conceptual model for all kind of health information about a person. This structure is called *medical biography*. Simply speaking we define for each medical concept like blood pressure or diabetes a number of dimensions, which capture all information of interest for this medical concept. For each person we can now define time series in these dimensions which mark all kinds of measurements or events for this person during its lifetime.

A schematic representation of one dimension of medical biographies is given in figure 2. S-data can be understood as projections of such dimensions of a medical biography, either for a fixed time, shown in figure 2 by the time \( t^* \), or as a projection for a specific age, shown in figure 2 by the symbol \( \Delta t \). EHR-data can be understood as information in a number of dimensions during a specific episode of care, marked in figure 2 as the gray shaded areas in the dimension of the first three persons.

Obtaining the datamart \( DM_S \) from a medical biography is usually not so difficult, because we can always identify dimensions which are rather close to the questions in a survey. Much more difficult seems the identification of the necessary EHR-data to build the datamart \( DM_P \), and to define the transformations and models necessary for making the two sources comparable. The basic steps for such an integration may be the following ones (\[5\]):

a) Identification of collectives of medical biographies which are of interest for the substantive question of interest (for example persons in a specific age group or persons with a specific disease).

b) Definition of the dimensions of the medical biographies, which contain the necessary information about the medical concept of interest.

c) Identification of possible EHR-data, which contain information for the dimensions of interest in the medical biographies.

d) Selection of EHR-data and transformation of the EHR-data onto the structure of the medical biography.

e) Definition of the necessary projections for the different dimensions of the biographies, which yield the summary attribute of the envisaged datamarts \( DM_P \) and \( DM_S \).

f) Calculations necessary for overcoming the differences according to different measurements in the dimensions and according to the sets \( A \) and \( B \) in figure 1.
References


[2.] BROWN N., REYNOLDS M., Strategy for production and maintenance of standards for interoperability within and between service departments and other healthcare domains. Short Strategic Study CEN/TC251/N00-047, CEN/TC251 Health Informatics, Brussels (Belgium), 2000.


