Interactive Visualization of Prescriptions of Drugs to Individuals within Large Populations—Analyses of Temporal Relationships of Events

Jimmy Johansson, Morten Andersen, Alexander Fridlund and Mikael Hoffmann

Abstract—This paper reports on work in progress on interactive visualization of prescriptions of drugs within large populations. A visualization prototype for interactive analysis is presented. The prototype has been developed in close collaboration between visualization researchers and domain experts within the area of pharmacoepidemiology. To illustrate the functionality of the prototype, data on treatment of hyperlipidemia with statins is used. The goal of the treatment with statins is to decrease the risk of future cardiovascular events among patients at an increased risk. Although no formal evaluation has been performed at this early stage of the development phase, feedback from end users has been positive. Many additional features have been proposed so there are several interesting and challenging directions for future work.

Index Terms—Interactive Visualization, Pharmacoepidemiology, Multivariate and Temporal Data

1 Introduction

During the last decade the volume of information gathered in large databases about health care has increased dramatically. The data can be generated from administrative health care records, electronic health records and quality-assurance tools, and even from claims databases. The accumulation of such large and complex data sets, combined with the possibility of record-linkage studies through the national personal identity number, creates many new opportunities for the growing field of pharmacoepidemiology, which studies the use, positive and negative effects of pharmaceuticals within populations of individuals to identify problems.

The goal of this paper is to introduce and discuss a novel way to interactive visualization of time-dependent multivariate data for large populations such as dispensings of prescribed drugs. More specifically, the paper presents work in progress and discusses the development of an initial prototype for interactive visualization of prescriptions of drugs to individuals within large populations. The prototype is the result of collaboration between visualization researchers and domain experts in the field of pharmacoepidemiology.

2 Background

Pharmacoepidemiology is the discipline of the research on the use of drugs and the effects of drugs in large numbers of people [7]. Both diseases caused by drugs (adverse drug reactions) and patterns of drug use (drug utilization research) are of interest. Specific challenges are the multiple categories of drugs and diagnoses, recurrent events, derivation of treatment duration from information on dosing, complex treatment sequences with intermittent drug use and switches, and the use of multiple simultaneous drugs. Established classifications of drugs (such as Anatomical Therapeutical Chemical Classification, ATC, established by WHO) and diagnoses (ICD) [3, 7] are important for analysing and disseminating the results internationally.

- *Jimmy Johansson is with C-Research, Linköping University, Sweden, e-mail: jimmy.johansson@liu.se.*
- *Morten Andersen is with the Centre for Pharmacoepidemiology, Karolinska Institutet, Sweden, e-mail: morten.andersen@ki.se*
- *Alexander Fridlund is with C-Research, Linköping University, Sweden, e-mail: alexander.fridlund@liu.se.*
- *Mikael Hoffmann is with the Department of Medical and Health Services, Linköping University, Sweden, e-mail: mikael.hoffmann@nepi.net.*

Manuscript received 31 March 2012; accepted 1 August 2012; posted online 14 October 2012; mailed on 5 October 2012.
For information on obtaining reprints of this article, please send e-mail to: tvcg@computer.org.

Since July 2005 all dispensings of prescribed drugs in Sweden (9.5 million inhabitants) have been gathered for each individual, and have been made available for research and for generating statistics by the Medical Board of Health and Welfare through the Swedish Prescribed Drug Registry. Similar databases exist in the other Nordic countries, as well as some other countries or regions, in order to facilitate research. Insurers or providers of health care also collect data sets with a similar structure.

Currently the access to data and the resources and time necessary for data management and analyses using database software, statistics applications and spreadsheets limits the use of these health data for exploratory analyses or rapid evaluation. Traditionally, data from different registers are retrieved and linked separately for each project, and conventional bar charts, histograms etc. are produced as part of the usual statistical analyses. Exploratory analyses follow the same path, and are therefore rarely done. Similarly, the use of interactive analyses that could add to the interpretation of patterns discovered in these data remains a huge challenge in these large multidimensional data sets.

Time-dependent, highly multivariate data, such as that exemplified by medical data of this nature, is one of the major challenges in data analysis [2]. Already involving the records of, potentially, millions of individuals, the data continue to expand with time, becoming ever larger and more difficult to manage but also, and more importantly, more difficult to represent in an understandable way. This creates new demands on tools and techniques to analyse and visualize the data sets. Traditional analysis methods suitable for hypothesis testing fall short as such tools, when applied to complex data, generates too many hypotheses that must be examined in the exploration process. Many excellent visual representations have been developed for high dimensional data [8] but the addition of the time factor renders most of them unhelpful since they are unable to show many time steps in a coherent way. Specific tools for visualization of temporal data exist [1] but need to be extended and customized to explore correlations and time-trends in order to support these new and important medical research and monitoring processes.

3 Interactive Visualization Prototype

This section describes the developed prototype and the case used in the examples.

The case considered is the treatment of hyperlipidemia with statins. The goal of treatment with statins is to decrease the risk of future cardiovascular events among patients at an increased risk. The use of statins has undergone major changes in the last years. The number of patients treated has increased, major patents have expired, and the reimbursement of statins has been revised in Sweden. The data set used for the pilot consists of all dispensings of statins to inhabitants from

40 to 85 years of age in Östergötland county during the period of 2007 through 2010. In addition the dispensings of any anti-diabetic agent during the period was also included as a proxy for diabetes mellitus treated with drugs. The data set was extracted as a part of a larger research project describing the impact of reimbursement changes during that period in Sweden. Data has been anonymized by replacing the unique patient identifier with a case number, as well as moving all birth-dates to the beginning of the corresponding year. Two different subsets of the database with 10,000 and 50,000 events respectively were used for the visualization.

The developed prototype consists of an interface based on coordinated and multiple linked views [6]. The primary visualization technique used is the Lexis diagram [5] which is a tool used in epidemiology to visualize relationships between events in time and a person's age. Figure 1 shows an illustration of the Lexis diagram where each point represents an event in time, in this case each event describes a drug prescription and each line connects all drug prescriptions made by a single individual over time.

After the data is loaded, the user can choose an arbitrary number of graphical views. Each view can be separately customized via a number of filters and visualization options. Standard interaction techniques such as zooming and panning is performed separately on each view. Figure 2 shows screen shots of possible configurations. Examples of interaction and visualization options are:

- down-scaling the data by sampling in order to focus the analysis on a specific subset.
- selecting data based on the Anatomical Therapeutic Chemical (ATC) Classification System for in-depth analysis of one or several substances, isolated or in relationships to each other.
- splitting the population based on sex in order to study differences between males and females.
- specifying a specific range in age to analyse differences or similarities between various age groups.
- selecting starting and end dates for analysis of a specific period in time.
- identifying patients that at some point in time have been dispensed an antidiabetic drug. The risk factor for cardiovascular disease is one of the conditions considered whether or not to treat with statins.
- colour coding the events and individuals based on ATC codes to analyse patterns in drug prescriptions over time.

The prototype has been developed in C++ and OpenGL to ensure efficient rendering. The interface is implemented using wxWidgets. The user interface as well as the specification of visualization and interaction techniques have been made in close collaborations with domain experts to ensure that the prototype supports as many tasks as possible and that the developed functionality is actually useful. Although no formal evaluation has been performed at this early stage of the development phase, feedback from end users has been positive. Many additional features have been asked for and there are many possible directions for future work, see discussion in section 4.

4 Conclusions and Future Work

This paper reports on work in progress on interactive visualization in the area of pharmacoepidemiology. A visualization prototype has been developed in close collaboration with domain experts within the area of pharmacoepidemiology. Preliminary results, supported by domain experts, suggests that the tool facilitates interactive analysis of this type of data and can be used for a number of different tasks that today are performed using traditional statistics software packages.

One area in which the developed prototype might be of particular value is in the identification of sequences and temporal relationships between dispensings of one or several different pharmaceuticals, which today is a time-consuming and cumbersome task. Possible confounders in the data-set such as changes in reimbursement systems, influence of media 'scares' and/or introduction in the market-place of

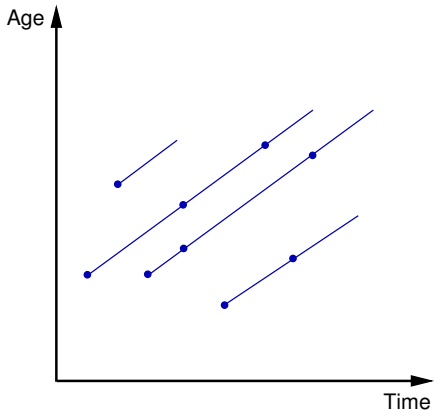


Fig. 1. A Lexis diagram is a representation of the relationship between events in time and a person's age. The x-axis shows calendar time and the y-axis shows the age of the patients. Each point represents an event in time, in this case each event describes a drug prescription. Each line connects all drug prescriptions made by a single individual over time.

substitute treatments within another drug class have to be recognized in order to be included in the analyse. The developed prototype might be used as a tool for pattern recognition in large temporal data sets in order to identify and handle such confounders.

Further development aims to visualize different measures important in pharmacoepidemiology. For one drug this include point- and period prevalence, incidence, adherence over time including overlap in dispensed amounts, persistence of treatment including duration of treatment and interruptions. For multiple drugs other measures such as combination of treatment, treatment sequence and interactions are important measures. In both cases time relationship to other important events such as diagnosis or in-patient treatment episodes are possible to analyse.

Another area of research is to enhance the existing Lexis diagram by, for example, make use of advanced blending schemes [4] in order to support efficient rendering of larger data sets. In addition, other visualization techniques, such as parallel coordinates, will be evaluated for use within this area.

Acknowledgments

This work was supported by the NEPI Foundation—Swedish Network for Pharmacoepidemiology, Norrköping Science Park, Sweden and the Swedish Research Council in the Linnaeus Centre CADICS.

References

[1] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski. Visual methods for analyzing time-oriented data. *IEEE Transactions on Visualization and Computer Graphics*, 14(1):47–60, 2008.

[2] C. Chen. Top 10 unsolved information visualization problems. *Computer Graphics and Applications*, 25(4):12–16, 2005.

[3] Guidelines for ATC classification and DDD assignment. WHO Collaborating Centre for Drug Statistics Methodology, Oslo, 2011.

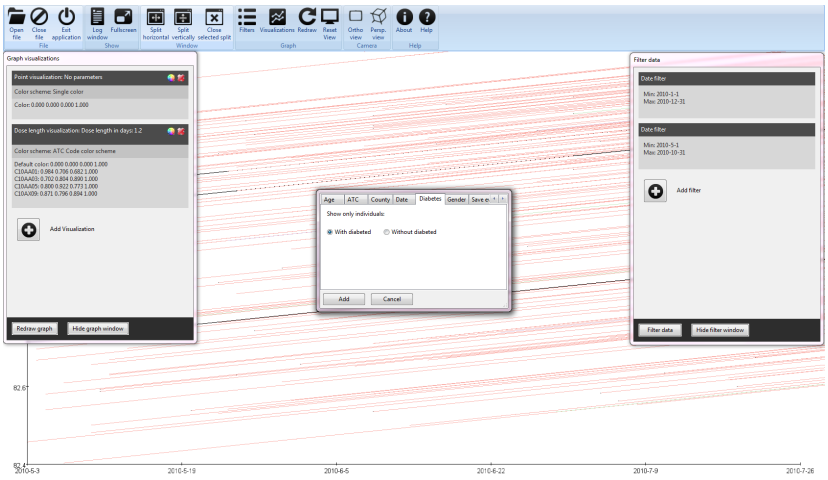
[4] J. Johansson, P. Ljung, M. Jern, and M. Cooper. Revealing structure within clustered parallel coordinates displays. In *Proceedings IEEE Symposium on Information Visualization 2005*, pages 125–132, 2005.

[5] N. Keiding. *Encyclopedia of Biostatistics*, volume 4, chapter Lexis Diagram, pages 2767–2769. Wiley, 2nd edition, 2001.

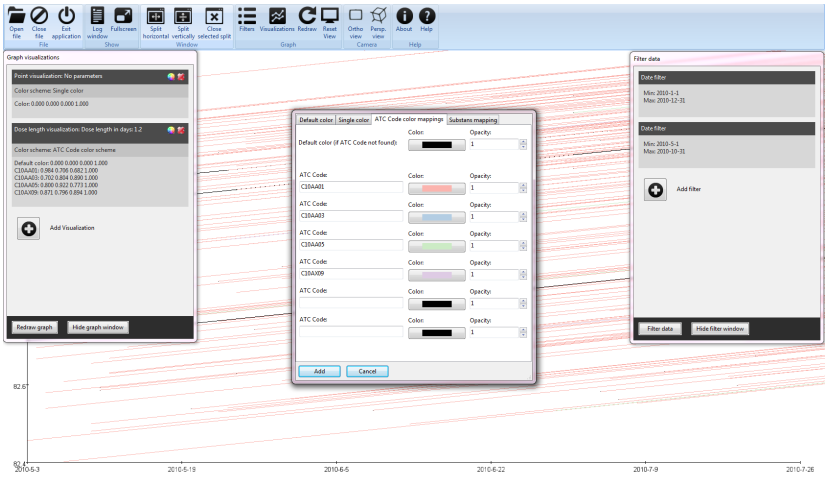
[6] C. North and B. Shneiderman. Snap-together visualization: a user interface for coordinating visualizations via relational schemata. In *Proceedings of the working conference on Advanced visual interfaces*, pages 128–135, 2000.

[7] B. L. Strom, S. E. Kimmel, and S. Hennesy, editors. *Pharmacoepidemiology*. John Wiley and Sons Ltd, 5th edition, 2012.

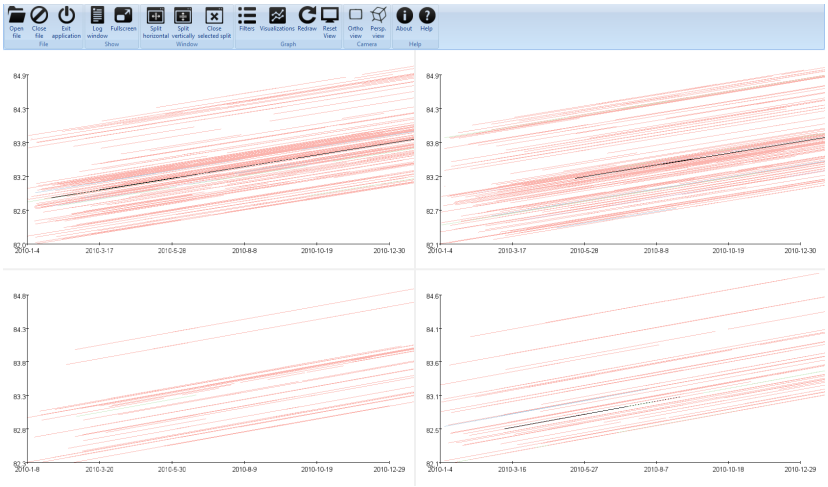
[8] P. C. Wong and R. D. Bergeron. 30 years of multidimensional multivariate visualization. In *Scientific Visualization, Overviews, Methodologies, and Techniques*, pages 3–33, 1997.



(a) Data filtering interface. This example shows filtering options enabling analysis of patients that at some point in time have been dispensed an antidiabetic drug.



(b) Visualization options interface. Data is colour coded based on the Anatomical Therapeutic Chemical (ATC) Classification System.



(c) An example of multiple views. Right to left: males and females. Top row: patients not treated with anti-diabetic drugs. Bottom row: patients treated with anti-diabetic drugs.

Fig. 2. Screen shots from the visualization prototype illustrating possible analyses of patterns in drug prescriptions over time using the Lexis diagram [5]. The data consists of patients with treatment of hyperlipidimia with statins.