A Medical Application to Bridge the Gap between Clinicians and Clinical Data

Mark FINKELSTEINa,1, Khawaja H BILALa and Michael PALESEa
aIcahn School of Medicine at Mount Sinai, New York, NY, USA

Abstract. There currently exists a gap between medical data and the clinicians who wish to use the data as the foundation for evidence-based medicine. The current disjointed workflow often requires statisticians to act as bridges in order to answer inquiries from clinicians. This can lead to suboptimal results as due to a gap between the fundamental understanding of clinical underpinnings and formal statistical interpretation of the data. To address this challenge, we have created a multi-platform application that allows clinicians to quickly and easily navigate and analyse the data. This application has been successfully implemented and tested within the Urology department at Mount Sinai Hospital for comprehensive analysis of complex urological data.

Keywords. Evidence-Based Medicine, Information Storage and Retrieval, Medical Informatics Computing, Clinical Research

Introduction

Since 1979, the Government of New York state has mandated that all health institutions, whether inpatient, outpatient or ambulatory centers, submit their medical records to a central database known as the Statewide Planning And Research Cooperative System (SPARCS).1 Previous studies have shown that Big Data is becoming vital in shaping and reforming the US Health Care system.2 Understanding the trends in medical practise has always been emphasized as a vital component to future planning and development.3 A recent example of this is the eradication of Prostate Specific Antigen testing from the urological outpatient setting while previously PSA testing was an integral part of routine. This occurred after the US Preventative Services Task Force studied health records of thousands of men and declared that the risk of prostate cancer was being over estimated.4 Studies show that a growing body of clinicians today are willing to analyse medical statistics and incorporate significant results into their practise.5 Several visual analytics systems have been developed in order to aid in the analysis of the increasing amounts of data being produced in all fields.6 These solutions tend to be generalized towards all types of data rather than individualized to the type of data. However, researchers have argued that Big Data’s role in healthcare is challenged by its particular complexity, security issues, privacy risks, as well as human skill. Furthermore, there has been limited work on Big-Data software written specifically for healthcare data.7 Our program aims to explore this area further and allows clinicians and researchers to easily access this wealth of data and output tables and statistical analysis for a virtually unlimited number

1 Corresponding Author. email: mark.finkelstein@mssm.edu
of queries. More importantly, it can be used to access the patient records of any medical specialty and can be adapted to suit any database. This program was tested at the Department of Urology at Mount Sinai Hospital, New York City for the study of trends and practices patterns of major urological surgeries over the past 30 years.

1. Methods

A multiplatform application to accommodate the wide variety of preferences in operating systems was created. Python, using the CPython implementation, was chosen to form the base of the application due to the ease with which it can be utilized on different operating systems. Hierarchical Data Format (HDF) files, accessed using the PyTables package, were chosen as the primary database for storing medical encounter data due to the ability to quickly compress, extract, and winnow data from these files. PyQt and Matplotlib were used for data representation and the Graphical User Interface (GUI). Pandas, NumPy, and SciPy packages were utilized to analyze and represent data.

2. Results

The application has been in consistent use on multiple operating systems for research purposes concurrent with its continued development since July 2014. The application allows users to quickly, easily, and flexibly set-up queries on the database (Figure 1). The user can then either export to CSV, which can be read by a number of programs such as Microsoft Excel, or to HDF, where the restricted dataset can be read more quickly or used for further analysis. The user can also make certain preset statistical calculations in order to better understand the data as well as to inform the user’s work with the data. These include univariate analysis, such as calculation of mean, standard deviation, kurtosis, and range for numeric variables and frequency for others such as ICD-9 variables. Also included is multivariate comparison, such as the ability to graph frequency against time, and the ability to create word clouds based on diagnosis or procedure codes.

3. Discussion

Preparing for the future of healthcare requires that the history of medical practice be taken into account. Observing the trends in urological surgeries has demonstrated that this program can definitely offer influential knowledge about the future of Urology. This insight can be used to alter physician practice and prepare for the oncoming changes. The program itself is malleable and thus can be applied to other fields and specialties. In conclusion, this program has the potential to allow researchers to easily utilize and answer questions pertaining to their field, by searching vast databases of patient records. The field of medicine is an ever-changing system and this program can evolve alongside with it.
Figure 1. A flowchart showing how a physician could perform a search to determine the diagnoses and ages of their patients. 

a. Files to load information from can be loaded in this view. In this way, a subset of the database can be exported and then loaded for querying and statistical analysis. This results in faster searches and more complex search options. 

b. This is the main view. From here, the user can choose to set up a query by moving fields of interests from the left table, which lists all the possible fields, to the right table. A criterion can be specified by double clicking on the field. 

c. Examples of sample output from the statistical view/generator, accessible from the main view.

References