Leveraging Domain Knowledge to Facilitate Visual Exploration of Large Population Datasets

William Hsu and Alex A.T. Bui
Medical Imaging Informatics Group
Dept of Radiological Sciences, University of California, Los Angeles

Abstract

Observational patient data provides an unprecedented opportunity to gleam new insights into diseases and assess patient quality of care, but a challenge lies in matching our ability to collect data with a comparable ability to understand and apply this information. Visual analytic techniques are promising as they permit the exploration and manipulation of complex datasets through a graphical user interface. Nevertheless, current visualization tools rely on users to manually configure which aspects of the dataset are shown and how they are presented. In this paper, we describe an approach that utilizes characteristics of the data and domain knowledge to assist users with summarizing the information space of a large population. We present a representation that captures contextual information about the data and constructs that operate on this information to tailor the data's presentation. We describe a use case of this approach in exploring a claims dataset of individuals with spinal dysraphism.

Introduction

On a daily basis, a large amount of data is generated as part of routine clinical patient care and captured as part of the electronic health record (EHR), disease registries, administrative databases, and other repositories. Retrospective analysis of the available data could yield new insights into diseases and provide evidence for improving best practices. For example, the data could support comparative effectiveness research (CER), which assesses the benefits, harms, and efficacy of different treatments for a disease based on patterns reported in a patient cohort [1]. However, working with observational data is often associated with challenges related to missing data, inherent and systematic errors, and ambiguity [2]. The use of information visualization and visual analytics to facilitate the analysis of sizeable, heterogeneous datasets are becoming popular because of their ease of use, leveraging our ability to interpret visual information quickly. In recent years, numerous visualization techniques have been developed to assist users with exploring high-dimensional datasets, identifying (causal) relationships and spatiotemporal patterns. Commercial packages such as Spotfire (TIBCO Software, Somerville, MA, USA) and Tableau (Tableau Software, Seattle, WA, USA) have created software packages that allow users to import data sources and specify the visualization and parameters with which the data is presented. While such visualization tools provide an intuitive and flexible means of generating visual representations from raw data, the task of specifying what data values to visualize and the attributes and behavior of the visualization are left to the user.

Arguably, current visualizations do not utilize contextual information about the data to influence how the information is presented. Here, context is defined as information that provides additional insight on how to correctly interpret a given data element. In particular, the utility of incorporating domain knowledge as part of the visualization process is underscored when considering increasingly higher dimensional data, as is the case when attempting to visualize large populations. First, a multitude of data elements (e.g., lab values, tumor burden, genetic variants) may be captured for each individual. To make trends within a group of patients interpretable, the dimensionality of the information needs to be reduced by abstracting or summarizing the information. However, this step is typically left to the user. Second, assumptions about the meaning of a data element and the interpretation of a given value may not be valid for a population, particularly if data has been drawn from different institutions. For instance, as results of genetic panels become captured as part of the patient record, direct comparison of gene expression values generated by differing next-generation technologies or results obtained using the same protocol but in subsequent batches may lead to erroneous conclusions. Hence, we argue that additional information beyond the value of the data element is important, and that a context-sensitive approach to visualization is necessary. We describe our initial attempts to define a common representation for capturing supplementary information about each data element and to demonstrate how the use of this domain knowledge facilitates users with navigating the large information space and interpreting patterns in a population.

In the following sections, we describe how our effort relates to existing visualization toolkits. We then discuss the design requirements of the toolkit, a representation for capturing context associated with the data, and modeling constructs for incorporating domain knowledge to automatically guide creation of an end visualization. Finally, we present a use case that implements the constructs to explore a large claims database of individuals with spinal dysraphism.

Background

Graphical toolkits such as prefuse [3], Piccolo [4], and IVTK [5] provide a generic framework for creating targeted information visualizations in a variety of complex domains. For instance, PatternFinder [6] is an application that leverages the Piccolo framework to permit visual search and comparison of temporal patterns across multiple (patient) histories. In medicine, a number of visualization tools have been created upon these frameworks to generate temporal, problem-oriented summaries of the patient record, perform case-based retrieval, conduct disease surveillance, and identify potential adverse drug events. While researchers performing CER typically utilize a combination of general visualization and statistical packages to identify patient cohorts and conduct their analyses, recent work has demonstrated the utility of incorporating targeted visualizations that facilitate CER within the context of the EHR. For example, VisualDecisionLinc [7] is a tool that visualizes information in the EHR to permit comparisons between an individual patient's case with trends identified from within a known population.

In general, the aforementioned visualizations span two ends of a spectrum: the generalized toolkits are the most flexible with respect to how data can be visualized, but heavily rely on users to specify how the dataset is presented. On the other hand, the targeted applications have strict data requirements and visualizations that are pre-configured to the specific task. Our approach attempts to bridge these two ends by providing a library of visualizations to present the heterogeneous nature of clinical data, yet tailoring these visualizations to reflect the semantic significance of the type and value of each data element. To do so, we have developed methods for characterizing the dataset and leveraging domain knowledge to select and configure the appropriate visualizations for a particular task.

Methods

Desiderata

In designing our context-sensitive visualization toolkit, the following criteria were considered:

- **Provide tailored views of the data**. Following the principle of adaptively selecting and configuring visualizations that best fit the available data and context, the population visualization should automatically generate a view that depicts relevant information at an appropriate level of abstraction based on the data. For instance, in visualizing an individual patient's lab values, a timeline of the measured values with related events (e.g., symptoms, prescribed medications, diseases) would be appropriate. However, the same visualization and level of detail cannot be used to present lab values for a hundred patients. A higher level of abstraction (e.g., heat map) may be more suitable.
- **Utilize domain knowledge.** The visualization should have awareness of the contextual information that is necessary to interpret the data. What is the semantic meaning of the reported value? How is it related to other data elements? Is the value within normal limits? What are the error bounds associated with the measurement? First, a representation is needed to annotate the raw data with this contextual information. Then, constructs are needed to encode this information as part of a visualization's attributes.
- **Convey temporal trends.** A common information need for individuals examining population datasets is providing tools for characterizing and comparing temporal patterns. For instance, visualizations should support comparison of the sequence of events between patients that have good outcomes versus those with poor outcomes. Techniques for aligning and summarizing different types of temporal data (e.g., time point, time interval) should be provided.

Data Representation

Contextual information plays a critical role in enabling users to understand and interpret the data being visualized. However, context is often lost or overlooked during the process of transforming raw data into its visual representation. We created a data model for representing the semantic meaning, properties, and observations of each data element. The basic entities in our model are loosely drawn from the phenomenon-centric data model described [8] and are summarized as follows:

- **Finding.** A finding is a broadly defined entity that identifies any source of observation or information. For example, a diagnostic test would be represented as an evidence entity with multiple observations (e.g., each time point that the panel was performed).

- **Observation.** An observation entity represents a value that has been measured or documented at a particular time point. The recorded value for an individual's creatinine level in the metabolic panel at a single time point represents an observation.
- **Attribute.** An attribute entity is associated with a finding or observation entity and captures the contextual information about that entity. Examples of attributes include data type (e.g., qualitative, quantitative), relation (e.g., greater than), units (e.g., Celsius), accuracy, precision, and certainty (e.g., exact, approximate).
- **State.** The state entity provides a higher order categorization for temporal stages that groups related findings and observations together for a single time point. For example, the state of a patient with renal failure would include relevant symptoms, labs, and treatments (i.e., routine dialysis).

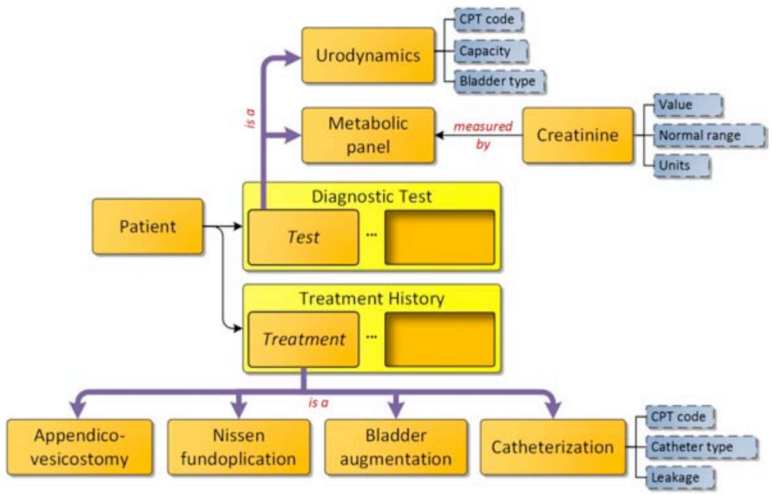

Figure 1: Schematic of the data representation showing findings (yellow), observations (orange), attributes (blue), relationships (black arrows), and hierarchy (bold purple arrows).

Components of the data model are summarized in Figure 1.

Basic Mapping

The visualization tool builds upon the five basic components of an information visualization described in [9]: data source, data set, visualization, view, and control. The data layer consists of the data source and data set, which contain mechanisms for loading and internally representing data from sources such as a delimited file or relational database. The visual layer consists of the visualization, view, and control components, which permit the transformation and interaction of the data through some graphical representation. For instance, a data set may be assigned a visual representation (e.g., circle, square) and attributes (e.g., size, position, color). This approach abstracts the data from its visual representation. Transforms are used to map the data layer to the visual layer.

Context-sensitive Constructs

The goal of these constructs is to permit visualizations to select and configure their appearance based on its underlying data and available domain knowledge. Concisely, we wish to exploit contextual information to anticipate the types of interactions that a user would perform on the data using the visualization. Borrowing from the seven categories of interaction described in [10], we describe methods that utilize context to influence how four of the interaction tasks are performed. These constructs are briefly elaborated below:

- **Connect.** The connect construct defines a relationship between two entities in a dataset. As depicted in Figure 2a, the construct can be used to link two semantically related entities: here, based on domain knowledge, we define a relationship between plasma fibrinogen values measured as part of a blood test and the usage of warfarin, a drug that is used to prevent blood clots, particularly in patients who suffered a heart attack. Defining this relation informs the visualization that measured fibrinogen values plotted over time (i.e., timeline) can be combined and overlaid with the visual representation of warfarin dosage (i.e., time bars). Connections are drawn from explicit relationships (e.g., directed arrows) encoded in the data representation.
- **Filter.** The filter construct allows visualizations to selectively incorporate and disregard certain data points based on a set of inclusion rules. A filter may be specified such that only abnormal laboratory values are included in the display, as shown in Figure 2b. Hence, if a patient has a lipid panel performed and triglyceride values fall within normal range, they would not be presented as part of the generated timeline. When dealing with many observations from a population cohort, the filter construct could be used to reduce the dimensionality of the dataset. Filters are defined based on attributes associated with entities in the data representation.

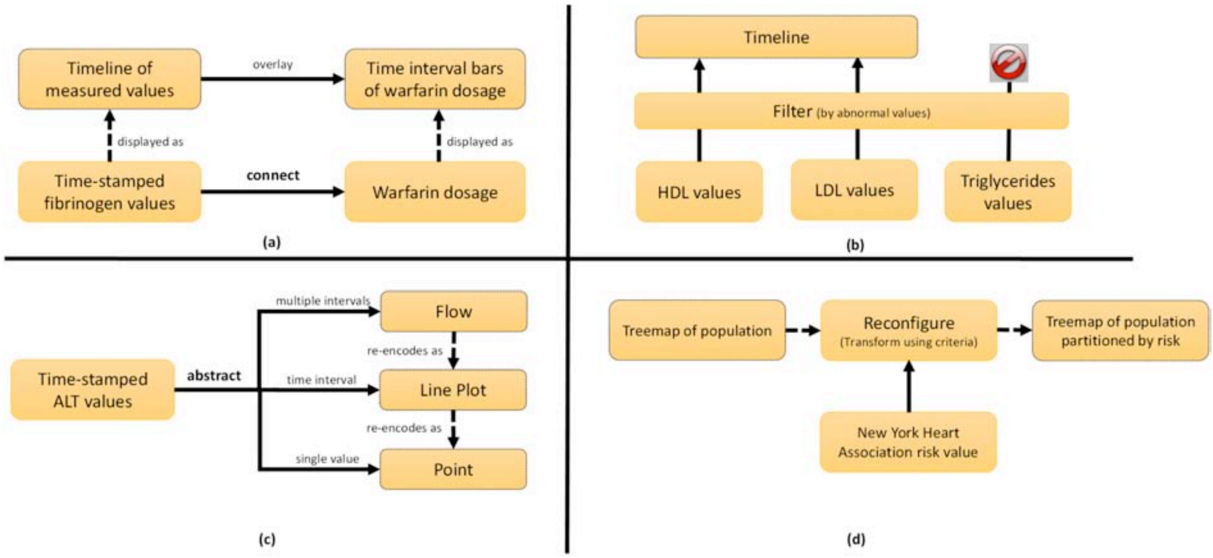

Figure 2: Constructs for manipulating graphical elements: (a) connect, which positions or layers semantically related information together; (b) filter, which removes information not relevant from the display; (c) abstract, which defines multiple layers of abstraction; and (d) reconfigure, which alters the visual presentation based on a transform function or pre-defined criteria.

- **Abstract.** The abstract construct allows data to be viewed at multiple levels of abstraction. Data can be abstracted in different ways depending on the type. For instance, Figure 2c demonstrates how time-dependent measurements (e.g., alanine aminotransferase (ALT) values) can be visualized in three different ways. A single value is visualized as a point. Multiple observations are plotted on a timeline. If measurements from multiple individuals need to be shown, a flow-based visualization is used, which does not present the actual values of the measurements but rather, the overall trend of the population. Furthermore, data can be abstracted hierarchically. In genomic data, the large number of measurements can be interpreted at the gene level (e.g., individual expression values), meta-gene level (e.g., genes grouped by similar function), or population level (e.g., significant gene expressions in a diseased population). By using domain information, the raw data can be abstracted and visualized in different ways, depending on the user's desired perspective. The data representation encodes hierarchical information as "is a" relationships between two entities.
- **Reconfigure.** The reconfigure construct provides a method for new insights derived from the data to be reflected in the visualization. Insights may be drawn from transformation functions that manipulate the data (e.g., compute the slope based on data points within a defined time window) or pre-defined criteria (e.g., rule-based classification). Figure 2d presents one example where a treemap depicting a hierarchical organization of the entire patient population (e.g., stratified by gender and age) can be reconfigured by adding a new criteria that stratifies the population further based on New York Heart Association (NYHA) risk values. Information on reconfiguring the visualization can be drawn from the relationships (e.g., rearrange the visualization based on a neighboring entity) and attributes (e.g., sort by value).

Implementation

Our visualization tool is built as a web-based platform using Data-Driven Documents (D3), a JavaScript library developed at the Stanford Visualization Lab [11]. D3 transforms a document object model (DOM) into visual representations based on the underlying data and provides a set of basic operations upon which elements of the document can be manipulated and transformed into a visual representation (drawn as a support vector graphic on the web page). An initial set of visualizations implemented in D3 has been incorporated: area charts, bar charts, line charts, pie chart, scatterplot, chord diagram, co-occurrence matrix, Sankey diagram, parallel sets, treemap, bubble chart, parallel coordinates, streamgraph, and choropleth. The data layer is implemented as a relational database. Transforms are implemented in Java in a Grails-based framework to manipulate objects from the database and generate the population-based visualization. Entities in the data representation (e.g., findings, observations, attributes) are encoded using eXtensible Markup Language (XML). The following section discusses our experiences in applying this framework to a specific patient cohort.

Results

We demonstrate the utility of our visualization tool in the context of performing an exploratory analysis on the hospitalization of myelomeningocele patients using administrative claims data from a cross-section of four states.

Clinical Background

Spinal dysraphism, which encompasses spina bifida and myelomeningocele, is a congenital condition that can result in a neurogenic bladder and pediatric urinary incontinence, and is considered one of the most prevalent permanently disabling birth defects. The Centers for Disease Control (CDC) estimates that 3 cases occur for every 10,000 births in the United States [12]. Approximately 180,000 Americans live with this condition [13], and given the advances in clinical care, many are now surviving into adulthood. But despite progress in the management of these patients, up to 40% develop some level of renal dysfunction over the course of their lives. Renal failure is the most frequent cause of mortality in this group [14, 15]. For urologists, the practical reality regarding the clinical management of these individuals is that it is an enormous challenge, in part due to the number and variety of interventions (e.g., catheterization, drugs, surgery) that may be performed over the course of the patient's life, the degree of patient compliance, the differences in practice between physicians, and the level of access to routine healthcare for follow-up. Moreover, it is not evident in many cases whether certain interventions are warranted and/or improve patient quality of life given the potential risk of complications and costs. The implications for improper care are severe, with the risk of long-term renal damage as high as 40% [16, 17]. Indeed, it remains unclear as to what clinical practices will balance control of neurogenic bladder versus quality of life, potential complications, costs, and ultimately, long-term outcomes. The goal of this case study is to utilize the population-based visualization to explore the dataset, demonstrating how domain knowledge can assist in identifying differences in care within the population.

Dataset

We performed the analysis using the Medicaid Analytic eXtract (MAX) file for 2008 from the Centers for Medicare & Medicaid Services (CMS), which covers eligible patients in four states spanning different geographical regions in the United States: California, Illinois, Texas, and New York. MAX is comprised of five files: Personal Summary File (PS), which contains demographics such as date of birth, gender, race; Other Therapy File (OT), which includes claim records for physician, imaging, and clinic services; Inpatient File (IP), which captures inpatient service information such as diagnoses, procedures, and length of stay; Long Term Care (LT) file, which lists services provided by nursing and intermediate care facilities; and Prescription Drug File (RX), which contains reimbursable drug claims.

Data Processing

The original data format (SAS file) was imported into a relational database that organized each MAX file as its own table linked together using the beneficiary identifier. Entries in the IP, OT, LT, and RX tables were associated with a time point/interval, diagnosis code (i.e., ICD-9), procedure code (i.e., CPT-4), provider type and location, and charged amount. To facilitate visualization of the data, we created additional tables in the database that integrate information from the above-mentioned tables to generate a master list of events (e.g., diagnosis, procedures, and drug prescriptions) ordered by service date. In total, 28,092 unique beneficiaries were identified in the dataset.

Encoding Domain Knowledge

To generate the data representation, we leveraged three sources of domain knowledge:

- **Existing knowledge sources.** We leverage existing biomedical ontologies and controlled vocabularies to provide the initial set of context for standardizing terms and identifying hierarchical and semantic relationships among terms. For example, the Medicaid dataset is primarily composed of diagnostic and procedure codes that need to be disambiguated. We leverage ontologies corresponding to the International Classification of Diseases (ICD-9) and Current Procedural Terminology (CPT-4) to provide information about the preferred term name, synonyms, semantic type, and concept unique identifier (e.g., to map to the Unified Medical Language System). We also utilized RxNORM to provide information about drugs, including their commercial name, dosage, administration route, and active ingredients.
- **Knowledge elicited from domain experts.** In consultation with a board certified urologist, we augmented the data representation to provide an additional layer of detail associated with the diagnostic tests and treatments pertinent to the urological management of patients with spinal dysraphism. The expert defined a list of entities and attributes corresponding to her information needs when reviewing a patient's record. For instance, entities

Table 1. Questions posed as part of the usability study and the types of constructs used to generate an answer.

Use Case	Construct(s)
Quantify the total number of Medicaid patients with spinal dysraphism treated in each requested state annually.	Abstract
Determine the proportion of subjects on renal dialysis.	Connect, Filter
Determine the number of hospitalizations and emergency room visits related to urinary tract infection during a 6-month period.	Connect, Filter, Abstract
Determine the average length of stay related to these hospitalizations	Reconfigure
Identify the proportion that undergoes renal ultrasounds within the first month, three months, and six months of life.	Connect, Filter
Determine the mean costs related to a hospitalization, including those related to emergency room admissions and elective surgery	Reconfigure

for kidney-related comorbidities such as renal mass, renal cyst, and hydronephrosis are defined along with attributes to provide characterization of the side (i.e., left, right, bilateral), grade, and severity (i.e., mild, moderate, severe). Each entity is associated with an attribute specifying the diagnostic or procedure code. Entities are also mapped to other biomedical ontologies or controlled vocabularies whenever possible. For example, urodynamics can be mapped to the standard term 'Urodynamic studies' in SNOMED-CT and classified as a diagnostic procedure. Using the taxonomy of SNOMED-CT, we can easily associate other instances of urodynamic studies under a single entity.

- **User intent.** We created several pre-defined tasks based on common types of queries that are of interest to the user. For example, comparison allows users to specify a set of criteria to divide the entire population into two or more subgroups. Within each subgroup, the user can define one or more variables to further stratify the population. If the user is only interested in comparing the prevalence of a single variable, pie charts depicting the proportion of patients that match the specified variable are generated for each subgroup. The user can then visually compare whether differences exist between subgroups. If multidimensional variables are selected, a more suitable visualization such as a heatmap or treemap is presented.

While defining a comprehensive data representation is a manual and potentially time consuming process, the encoded knowledge can be re-used or extended as part of subsequent studies. We also have developed programmatic tools to help translate existing ontologies encoded in web ontology language (OWL) to our XML schema so that they may provide an initial source of context. We also acknowledge that the MAX dataset does not provide specific details about a patient's diagnosis or procedure, but we anticipate utilizing the visualization tool to explore structured patient records seen at our institution's urology clinic. The data representation can also be utilized to inform the information extraction task.

Use Case

The visualization tool guides the user with exploring the dataset by providing an initial set of filters that allow users to identify subsets of the population. In the case of the MAX dataset, the user initially has access to over 28,000 individuals who meet the inclusion criteria. Table 1 provides a list of questions that drive potential use cases for utilizing the developed representation and constructs to generate a context-sensitive display. Generally, interaction with the tool proceeds as follows. Once the data and contextual information have been loaded into the tool, the user interacts with a set of filters to specify his/her information needs by defining the task (i.e., description) and identifying the relevant variables (i.e., drawn from the database). Once selected, one or more visualizations are automatically presented and configured based on the number and type of variable selected. A specific example of this process is illustrated to obtain descriptive statistics of hospitalizations related to urological disorders within the spinal dysraphism population:

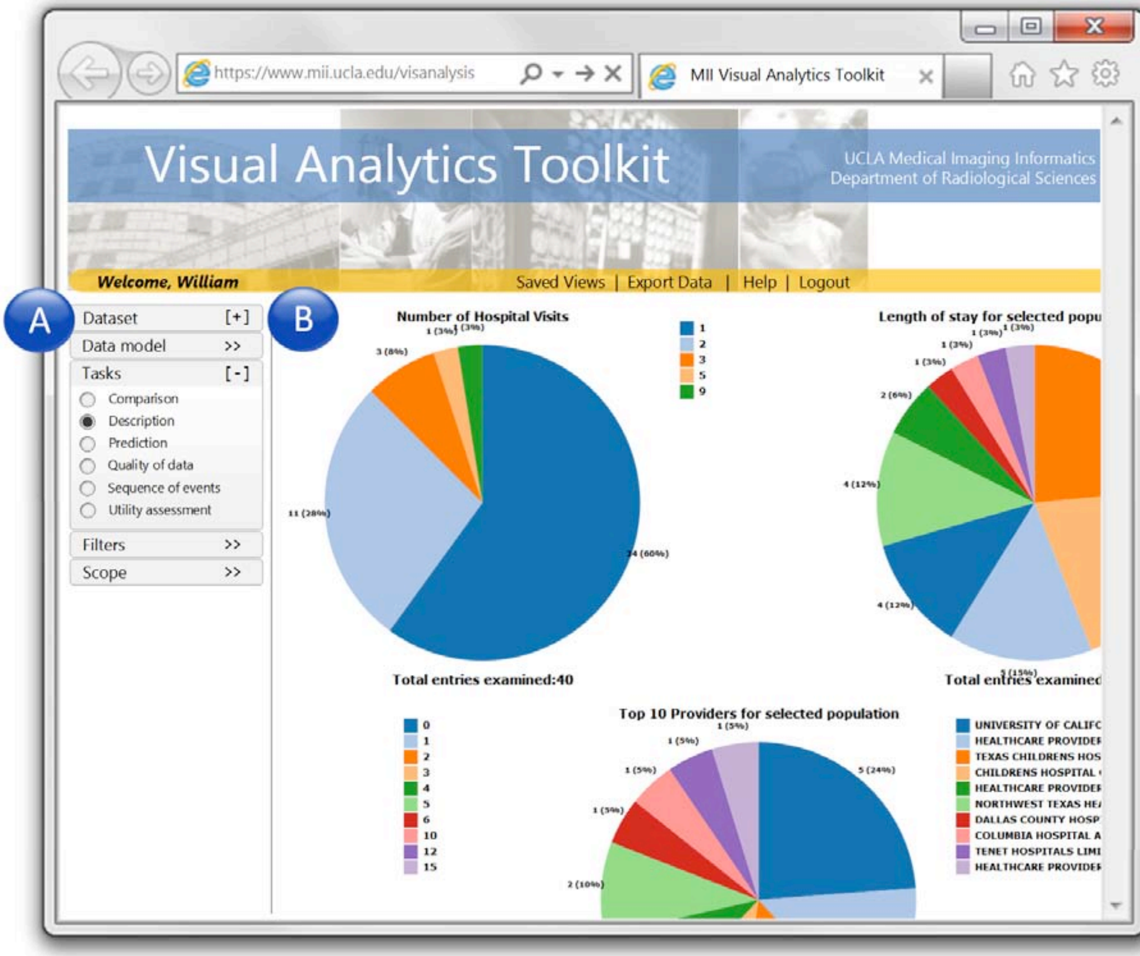

Figure 3: Depiction of the user interface demonstrating how the population dataset can be visualized to understand urologic outcomes: (a) a set of filters provides users with a (tailored) list of options based on task and data available. (b) Once filters have been selected, the selected variables are visually presented guided by the data representation.

1. The user first loads the dataset and data representation. Here, the representation explicitly defines which data elements in the database are related to hospitalization, including start/end dates, provider information, and associated diagnostic and procedure codes. In addition, coded values used to represent each data element are mapped to existing ontologies and knowledge sources. This information provides the means to identify individuals who have renal failure (e.g., ICD-9 code 584) or end-stage dialysis (e.g., CPT-4 code 90960), which represent urological outcomes of interest.

2. Next, a pre-defined task is selected. Each task is associated with a set of visualizations and transformations. For example, the 'sequence of events' task is used to compare the temporal ordering of events between two patient subpopulations. This task is associated with a transformation that aligns patient records based on a user-defined sentinel event; a Sankey-like diagram is then used to visualize the results. Such task would be useful for identifying temporal similarities of procedure codes for individuals who have achieved better urologic outcomes (e.g., individuals who choose to have catheterization procedure done earlier in life). In this example, the user has selected the 'description' task, which generates a summary view of relevant characteristics.

3. Filters can be specified to narrow what subset of the population is included in the analysis. Users may determine which data elements may be included in or excluded from the analysis. In addition, propositional logic is used to define constraints for each data element (e.g., ranges of possible values). Here, we define our population to

only include individuals who have had end-stage renal disease and have been hospitalized at least once by specifying a range of ICD-9 and CPT-4 codes that match that criteria.

4. The amount of detail shown to the user is determined by specifying a scope. A scope may be defined based on a temporal constraint (e.g., 1-month versus 1-year) or a hierarchical constraint (e.g., aggregation of concepts based on shared semantic type). In this use case, we specify the temporal constraint to be days hospitalized.
5. Based on the selected criteria, the visualization tool utilizes the available constructs and contextual information to generate an initial display. The resulting visualization is depicted in Figure 3: the user is presented with pie charts depicting the number of hospital visits, the length of stay, and the top providers for the matching population. Depending on user-selected data types and constraints, other visualizations may be leveraged as well. Users can interact with the interface to override the selected visualization (e.g., display the data using a bar chart rather than a pie chart) based on their preferences.

Discussion

Visual analytic techniques provide researchers and clinicians the ability to quickly explore and understand large quantities of information by interacting with a visual representation of the complex dataset. For complex domains that handle large, high-dimensional datasets, tools that assist users with generating the appropriate views of the data would facilitate tasks such as hypothesis generation and CER studies. In this work, we describe modeling and visualization constructs for tailoring the presentation of population datasets. We demonstrate the utility of constructs that utilize contextual information to connect, filter, abstract, and reconfigure data of individual patients to identify trends in a population. Given the increasing amount of observational data being routinely captured in healthcare, interfaces used to explore and manipulate this information need the ability to adapt the information presented to users: not all information is equally important. In this work, we explore the need to consider not only the type of data being visualized (e.g., numerical, categorical), but also the value and semantic meaning of the data. Providing contextual information along with the data is critical to ensure that the resulting visualization correctly conveys the intrinsic assumptions and limitations associated with the data.

While this work highlights four context-sensitive constructs, other constructs may be identified as the paradigm is used in different applications. To assess the utility and scope of the constructs, we are undertaking an evaluation to assess the usability of the population-based visualization tool and the effectiveness of utilizing domain knowledge to tailor the presentation of information. We have recruited a convenience sample of five participants to participate in an internal evaluation study that includes medical students, attending physicians, and health services researchers. Participants are first briefly introduced to the visualization tool and guided through the process of answering one sample question. Then, each participant is tasked with answering a set of twenty questions based on information from the MAX dataset. Half of the questions are to be answered using the visualization tool where the user must define the variables and configure the visualizations manually. The remainder of the questions is to be answered using the tool where the context-sensitive constructs are enabled, and the visualizations are automatically configured. After the study, participants are asked to complete a brief survey that elicits ratings about their experience with the tool using a Likert scale between 1 (worst) to 5 (best). The participants would also be asked whether they would suggest additional constructs that would facilitate the exploratory process. Based on the initial results of the internal study, we will estimate the necessary sample size to detect a difference in time required to complete each question and accuracy of the answer between the two methods with a power of 90%.

Our efforts thus far have focused on visualizing claims data that are limited in amount of detail and heterogeneity. Other sources such as the EHR and high-throughput genomic sequencing exemplify datasets with greater variability and complexity. As part of future work, we intend to review the full records of patients with spinal dysraphism seen at our institution, enumerating the different types and values reported and mapping them to appropriate visualizations. In addition, while the visualization tool described as part of the use cases was designed based on regular feedback from our clinical collaborators, we hope to characterize how easily target users can understand the data as presented through the visualizations. Results of the aforementioned usability study will help elucidate whether differences exist in the amount of time required to identify patterns and the quality of insights gleaned from the dataset when utilizing the context-sensitive visualization versus a basic set of visualizations.

Acknowledgements

The authors would like to thank Drs. Bernard Churchill, Mark Litwin, and Jenny Yiee for their insights and support of this project. Departmental funds from Department of Radiological Sciences and Urology at UCLA were used to support this work.

References

1. Sox HC, Greenfield S (2009) Comparative Effectiveness Research: A Report From the Institute of Medicine. Ann. Intern. Med., 151(3):203-205.
2. Hersh WR, Weiner MG, Embi PJ, Logan JR, Payne PR, Bernstam EV, Lehmann HP, Hripcsak G, Hartzog TH, Cimino JJ, Saltz JH (2013) Caveats for the use of operational electronic health record data in comparative effectiveness research. Medical care, 51(8 Suppl 3):S30-37.
3. Heer J, Card SK, Landay JA (2005) Prefuse: a toolkit for interactive information visualization. Proceedings of the SIGCHI conference on Human factors in computing systems. ACM, pp 421-430.
4. Bederson BB, Grosjean J, Meyer J (2004) Toolkit design for interactive structured graphics. Software Engineering, IEEE Transactions on, 30(8):535-546.
5. Fekete JD (2004) The infovis toolkit. Information Visualization, 2004. INFOVIS 2004. IEEE Symposium on. IEEE, pp 167-174.
6. Fails JA, Karlson A, Shahamat L, Shneiderman B (2006) A Visual Interface for Multivariate Temporal Data: Finding Patterns of Events across Multiple Histories. Visual Analytics Science And Technology, 2006 IEEE Symposium On, pp 167-174.
7. Mane KK, Bizon C, Schmitt C, Owen P, Burchett B, Pietrobon R, Gersing K (2012) VisualDecisionLinc: A visual analytics approach for comparative effectiveness-based clinical decision support in psychiatry. J Biomed Inform, 45(1):101-106.
8. Bui AAT, Aberle DR, Hooshang K (2007) TimeLine: Visualizing Integrated Patient Records. IEEE Trans Inf Technol Biomed, 11(4):462-473.
9. Heer J, Agrawala M (2006) Software design patterns for information visualization. IEEE Trans Vis Comput Graph, 12(5):853-860.
10. Yi JS, ah Kang Y, Stasko JT, Jacko JA (2007) Toward a deeper understanding of the role of interaction in information visualization. Visualization and Computer Graphics, IEEE Transactions on, 13(6):1224-1231.
11. Bostock M, Ogievetsky V, Heer J (2011) D3: Data-Driven Documents. IEEE Trans Vis Comput Graph, 17(12):2301-2309.
12. Shin M, Besser LM, Siffel C, Kucik JE, Shaw GM, Lu C, Correa A (2010) Prevalence of spina bifida among children and adolescents in 10 regions in the United States. Pediatrics, 126(2):274-279.
13. Danielson M, Thibadeau J, Devine O, Correa I (2009) Population estimate of people living with spina bifida in the United States, 2004. First World Congress on Spina Bifida Research and Care, Orlando, FL.
14. Smith K, Mizokawa S, Neville-Jan A, Macias K (2010) Urinary continence across the life course. Pediatr Clin North Am, 57(4):997-1011.
15. Bowman RM, McLone DG, Grant JA, Tomita T, Ito JA (2001) Spina bifida outcome: A 25-year prospective. Pediatr Neurosurg, 34(3):114-120.
16. Lewis M, Webb N, Stellman-Ward G, Bannister C (1994) Investigative techniques and renal parenchymal damage in children with spina bifida. Eur J Pediatr Surg, 4S1:29–31.
17. Muller T, Arbeiter K, Aufricht C (2002) Renal function in myelomeningocele: Risk factors, chronic renal failure, renal replacement therapy and transplantation. Curr Opin Urol, 12:479–484.