

Visualisation of Integrated Patient-Centric Data as Pathways: Enhancing Electronic Medical Records in Clinical Practice

Joao H Bettencourt-Silva¹ Gurdeep S Mannu² Beatriz de la Iglesia³

¹ Clinical Informatics, Department of Medicine,
University of Cambridge, CB2 0QQ, United Kingdom
jhb56@medschl.cam.ac.uk

² Nuffield Department of Population Health,
University of Oxford, OX3 7LF, United Kingdom
gurdeep.mannu@gtc.ox.ac.uk

³ School of Computing Sciences,
University of East Anglia, Norwich, NR4 7TJ, United Kingdom
b.iglesia@uea.ac.uk

Abstract. Routinely collected data in hospital Electronic Medical Records (EMR) is rich and abundant but often not linked or analysed for purposes other than direct patient care. We have created a methodology to integrate patient-centric data from different EMR systems into clinical pathways that represent the history of all patient interactions with the hospital during the course of a disease and beyond. In this paper, the literature in the area of data visualisation in healthcare is reviewed and a method for visualising the journeys that patients take through care is discussed. Examples of the hidden knowledge that could be discovered using this approach are explored and the main application areas of visualisation tools are identified. This paper also highlights the challenges of collecting and analysing such data and making the visualisations extensively used in the medical domain.

This paper starts by presenting the state-of-the-art in visualisation of clinical and other health related data. Then, it describes an example clinical problem and discusses the visualisation tools and techniques created for the utilisation of these data by clinicians and researchers. Finally, we look at the open problems in this area of research and discuss future challenges.

Keywords: Visualisation, Big Data, Clinical Pathways, Data Mining, Knowledge Discovery, Data Quality, Decision Making, Medical Informatics

1 Introduction

Hospitals routinely collect data related to the interaction of patients with different departments and medical specialties. Traditionally this information was

recorded in paper notes yet more recently there has been an increasing shift towards the adoption of electronic medical records, as the statistics from the Electronic Medical Record Adoption Model (EMRAM) demonstrate (<http://himss.eu/emram>), yet in many cases, researchers may still need to collate information manually [1] and methodologies to facilitate this process are relatively unexplored [2]. Clinical data is typically complex and may pertain to diagnoses, admissions and discharges, prescriptions, treatments, biomarkers and blood tests, outcomes and other clinical findings. As a result, patients leave footprints on many hospital systems, but such prints are not often connected to provide a pathway indicative of their journey through care, nor are they presented at the aggregated level. In the context of important diseases such as cancer or stroke, the journey of patients from diagnosis to outcome would provide a unique perspective that could aid clinicians to better understand disease processes and provide valuable information on optimal treatment. Hence, an initial challenge is to gather data from multiple EMR systems and construct meaningful data structures that can encompass all of the relevant information pertaining to a given patient and a given disease over time. We have named such data structures *clinical pathways* and have provided a methodology to build them [2,3]. Note that some researchers refer to clinical pathways as the standardised and normalised therapy pattern recommended for a particular disease [4]. Other researchers have focused on mining common pathways that show typical disease progression based on hierarchical clustering and Markov chains [5]. Our pathways relate to the journey followed by the patient through care and they may align with the recommended guidelines for a particular disease but may also deviate from it.

Visualisations of pathways, at the individual or aggregate level, when well presented and of high quality, could help clinicians to interact with such data and give them a view of patients and disease progression that was otherwise hidden away in databases. This would enable them to utilise the power of the big data in their environment, a very topical subject which currently holds much promise. For example, Shneiderman et al. [6] state that “while clinical trials remain the work horse of clinical research there is now a shift toward the use of existing clinical data for discovery research, leading researchers to analyse large warehouses of patient histories”. The visualisation of this big data is a critical topic and the specific subject of this paper.

In the context of medical data mining, clinical pathways, as we define them, require consistent pre-processing techniques, innovative data mining methods and powerful and interactive visualisation techniques. They also present the challenges of data privacy which has to always be maintained when dealing with patients’ data. We discuss some of these challenges and present some solutions in this paper, particularly focusing on the visualisation aspects.

This paper is organized as follows: to ensure a common understanding we provide a short glossary in section 2; we examine work on visualisation of medical data that is relevant in the context of the problem we present in section 3; we then provide some background information about clinical pathways, their

construction, their visualisation and the challenges of such an approach in section 4. We then discuss the processes of visualization of aggregated pathways in section 5 and their areas of application in section 6. Finally, we discuss problems in the field and conclude with prospects for the future.

2 Glossary and Key Terms

Electronic Medical Record (EMR): can be characterised as “the complete set of information that resides in electronic form and is related to the past, present and future health status or health care provided to a subject of care” [7].

Medical Informatics: is the interdisciplinary study of the design, development, adoption and application of IT-based innovations in healthcare services delivery, management and planning [8]. Medical informatics is also called health care informatics, health informatics, nursing informatics, clinical informatics, or biomedical informatics.

Data Mining: is an analytic process designed to explore large amounts of data in search of consistent patterns and/or systematic relationships between variables, and then to validate the findings by applying the detected patterns to new subsets of data [9].

Medical Patterns: these are frequently appearing sequences of treatments, diagnoses, etc, that are associated with unusually positive or negative outcomes [10].

Visual Analytics: denotes the science of analytical reasoning facilitated by visual interactive interfaces [11].

Data Quality: includes (physical) quality parameters such as: Accuracy, Completeness, Update status, Relevance, Consistency, Reliability and Accessibility [12].

Clinical Pathway: in the context of this paper it is defined as an ordered set of patient-centric events and information relevant to a particular clinical condition [3]. It can be considered as a suitable data structure for routine data extracted from EMRs that records the actual journey of the patient for a given condition. Others have defined it as “a map of the process involved in managing a common clinical condition or situation” [13]. Hence in the second definition the clinical pathway may embody the *ideal* or recommended pathway and enumerate regular medical behaviours that are expected to occur in patient care journeys and may, therefore, serve as a checkpoint for the performance of the actual pathway.

Temporal abstraction: this refers to the task of creating interval-based concepts

or abstractions from time-stamped raw data. In the context of electronic clinical data, data summaries of time-oriented data can help for example when physicians are scanning a long patient record for meaningful trends [14].

Clinical guidelines: are systematically developed statements designed to help practitioners and patients decide on appropriate healthcare for specific clinical conditions and/or circumstances [15]. They may articulate a desired clinical pathway.

3 State-of-the-Art

One of the main characteristics of clinical data is its temporal nature. EMRs are composed of longitudinal event sequences which can sometimes be a concurrent set of treatments for various conditions undertaken by a patient over time. Another important characteristic is the complexity of the data, which can include many different data types, support many levels of granularity and is associated with extensive domain knowledge that may be required for context. Additionally, the type of analysis we want to support may require techniques that take into account individual patients, or aggregate at the cohort level. As we are focusing on visualisation, we need to generate visual user interfaces that can represent such complexity efficiently and effectively without overwhelming the user. We need to provide query engines and mining methods that can deal with the temporal and complex nature of the data with efficient interactions. We also need to ensure that the systems produced are evaluated effectively, which is difficult when evaluation requires the involvement of busy medical practitioners. In this section, we review how researchers have tackled some of these problems so far.

As a starting point, reviews and surveys on the subject of visualisation of EMR data provide a good introduction to this topic. Turkay et al. [16] give a recent introduction to the visualisation of large biomedical heterogeneous data sets and point out the need for mechanisms to improve the interpretability and usability of interactive visual analyses. They also stress the challenge of integrating data from additional sources, such as the “microscopic” world (systems biology), the “omics” world or the “macroscopic” (public health informatics) world, as we move towards precision medicine.

Rind et al. [17] provide a survey comparing a number of state-of-the-art visualisation research systems for EMR, and separately give examples of visualisations produced by commercial systems. They also give a summary of other reviews of this subject. Roque et al. [18] also give comparisons of the key information visualisation systems for clinical data. Similarly, West et al. [19] provide a systematic survey of works between 1996 and 2013. Their article is part of a special issue dedicated to visual analytics to support the analysis of complex clinical data [20]. Lesselroth and Pieczkiewicz [21] discuss a number of strategies for visualising EMRs. More generically, methods for visualising time oriented data have also been surveyed [22].

Time oriented clinical data has been considered to be important by a number of researchers. Early work on visualisation of personal histories [23] produced a system called Lifelines that used graphical time scales to produce a timeline of a single patient’s temporal events. Medical conditions could be displayed as horizontal lines, while icons indicated discrete events, such as physician consultations. Line colour and thickness were used to illustrate relationships or the significance of events. Application of Lifelines to medical records was further explored in [24]. Lifelines is the basis for many other systems that visualise time oriented clinical data. The evolution of Lifelines produced a system called Lifelines2 [25] that displays multiple patient histories aligned on sentinel events to enable medical researchers to spot precursor, co-occurring, and after-effect events.

Further work by the same team resulted in LifeFlow [26], which presents a prototype for the visualisation of event sequences involving millions of patients. LifeFlow was one of the first systems to provide an overview and enable the answering of questions such as “what are the most common transfer patterns between services within the hospital”. Hence Lifeflow attempts to summarise all possible sequences, together with the temporal spacing of events within the sequences. It provides one visual abstraction that represents multiple timelines so it addresses the problem of aggregation. In terms of the interaction capability, which has become a key issue in visualising clinical information, LifeFlow [26] provides zooming, sorting, filtering and enables further exploration of events by hovering the cursor over parts of the visualisation. It also enables the user to select non-temporal attributes as the basis for aggregation. This enables comparison between different groups.

Shahar et al. [14] also worked with temporal clinical data. In particular they discuss the extraction of temporal abstractions from electronic data. Such temporal abstractions combine a domain knowledge-base with interval-based concepts. A quoted example is the abstraction of Bone Marrow toxicity from raw individual hematological data. The domain knowledge in this case would establish the context such as following Bone Marrow Transplantation using a particular therapy protocol. A simpler abstraction may be fever from multiple measures of raised temperature over time. Temporal abstractions can support intelligent decision-support systems or be used for the monitoring of clinical guidelines. However, Shahar et al. argue that temporal abstractions can only be truly useful in a clinical setting if they are accompanied by interactive visualisation and exploration capabilities which can also take into account medical domain knowledge. For this, they developed a system called KNAVE-II, a development of a previous system [27]. The work does not provide, however, capabilities for aggregation of patients according to some dynamic criteria. In further work [28], the authors provided such capability under a system called VISITORS.

The issue of introducing context when evaluating patterns in a clinical setting is also important in other scenarios. For example, Duke et al. [29] present a system for incorporating knowledge such as a patient’s relevant co-morbidities

and risk factors when evaluating drug-drug interactions to improve the specificity of alerts.

Analysis based on comparison of cohorts is also prevalent. Huang et al. [30] describe a system for exploratory data analysis through a visual interactive environment to show disease-disease associations over time. The system simplifies visual complexity by aggregating records over time, clustering patients and filtering association between cohorts. The main visualisation methods used to study disease trajectories over time are Sankey diagrams [31].

Wong et al. [32] proposed INVISQUE, an interactive visualisation to support both medical diagnosis and information analysis and discussed the key issues that need to be addressed when designing interactive visualisation systems for such purposes. CareVis [33] is another system, specifically designed to provide visualisation of medical treatment plans and patient data, including contextual information on treatment steps. It utilises a language called Asbru, designed to represent clinical guidelines and protocols in eXtensible Markup Language (XML). Challenges of the data include hierarchical decomposition, flexible execution order, non-uniform element types and state characteristics of conditions. CareVis utilises multiple integrated views [34] to represent logical and temporal aspects of the treatment data. The views can be coupled with colour, brushing and navigation propagation, hence elements in one view can be linked to the same elements in the other views allowing for interaction with the visualisation.

Another recent work using Asbru, following from CareVis, and specifically designed to analyse compliance with clinical guidelines is presented by Bodesinsky et al. [35]. The authors use visualisation to integrate information about executed treatments with Computer Interpretable Guidelines. Combining views from observation, treatment and guidelines is becoming increasingly important in the clinical setting.

Very recent work on visualisation of temporal queries, which enables clinicians to extract cohorts of patients given temporal constraints is presented by Krause et al. [36]. Retrospective cohort extraction in the traditional way involves a long and complex process and requires involvement from doctors and SQL query specialists. SQL queries do not cater well for temporal constraints and query engines may not optimise well such queries, making the process difficult and inefficient. A system called *COQUITO* is proposed as a visual interface for building COhort QUeries with an IIterative Overview for specifying temporal constraints on databases. The query mechanism is implemented by a visual query user interface and provides real-time feedback about result sets. It also claims to be backed by a Temporal Query Server optimized to support complex temporal queries on large databases. Another system for constructing visual temporal queries is DecisionFlow [37]. DecisionFlow enables interactive queries on high-dimensional datasets (i.e. with thousands of event types).

Given the amount of complex data that needs to be visualised in the context of medical systems, one common problem is the dense display that can result and the difficulty this represents for the user. For example, Kamsu-Foguen et al. [38] discuss the need for intelligent monitoring systems that can help users

with the massive information influx. This may require the capturing of domain knowledge to form a physiological/process model as part of the expert interface. It may also require the use of machine learning to improve interaction of machines and humans (e.g. reducing data input by inducing entries based on previous interactions). The software proposed can integrate visual and analytical methods to filter, display, label and highlight relevant medical information from patient-time oriented data. At the same time, it can learn from interactions between medical staff and the system in a particular context, such as modification of a prescription. It could then be used for instance to capture domain expert knowledge in respect to medical guideline compliance.

An issue that is also now receiving attention is the efficiency of visual analytic algorithms as dataset grows. According to Stolper et al. [39] “in the context of medical data, it is common to find datasets with tens of thousand of distinct type of medical events, thousands or even millions of patients and multiple years of medical data per patient.” There are typically delays in the workflow of analysts launching queries, inspecting results, refining queries and adjusting parameters and relaunching queries. In this scenario, Stolper et al. propose the use of progressive visual analytics that enable analysts to explore meaningful partial results of an algorithm as they become available and interact with the algorithm to prioritise subspaces of interest. The interface also enables the user to adjust parameters as algorithms are running, re-start the running but also store results obtained until that point so that the user can resume previous run if required.

There are parallels between information visualisation and data mining [40]. Visual Data Mining can integrate the human in the data exploration process and can be seen as a hypothesis generation process based on visualisations [41]. Data Mining analysis is also being applied to clinical data in conjunction with visualisation techniques in order to extract knowledge, for example by identifying outliers and deviations in health care data [42]. For clinical pathways, pathway mining is also prominent and often associated with process mining using clinical workflow logs to discover medical behaviour and patterns [4]. Perer and Wang [10] have integrated frequent pattern mining and visualisation so that the resulting algorithms can handle multiple-levels of detail, temporal context, concurrency and outcome analysis and visualise the resulting frequent event sequences from EMR. This has resulted in a prototype system, Care Pathway Explorer [43], which can correlate medical events such as diagnosis and treatments with patient outcome. The system has a user-centric visual interface which can represent the most frequent patterns mined as bubbles, with the size corresponding to number of times a particular event occurs. It also uses Flow Visualisation to see how the bubbles connect to each other.

Measuring the quality of the data to be used is an important issue, as routinely collected data can be of variable quality. It would be very useful for any system that works with EMR to provide some quality measurements that can be used for the purposes of including or excluding records for further queries and clinical studies. For example, Tate et al. [44] elude to work in this area as part of

their attempt to construct a system that enables querying of large primary care databases to select GP practices for clinical trials based on suitability of patient base and measures of data quality.

Another important topic is the visualisation of biological and “omics” data [16]. In systems biology, Jeanquartier et al. [45] carried out a large survey of databases that enable the visual analysis of protein networks. Systems such as the NAViGaTOR 3 extend the basic concept of network visualisation to visual data mining and allow the creation of integrated networks by combining metabolic pathways, protein-protein interactions, and drug-target data [41]. Other techniques, such as multilevel glyphs, have been proposed as a multi-dimensional way to visualise and analyse large biomedical datasets [46] and there is still a high demand for specialized and highly integrative visual analytic approaches in the biomedical domain [40], particularly as we move towards personalised medicine.

The evaluation of information visualisation tools is one of the open challenges in this area. Often carried out by controlled experiments and the production of usability reports, this are however described by Shneiderman and Plaisant [47] as helpful but falling short of expectations. They describe a new paradigm for evaluation in the form of Multi-dimensional In-depth Long-term Case studies (MILCs) that may begin with careful steps to gain entry, permission and participation of subjects and be followed by intense discussions which provide key data for evaluations. As MILCs provide multiple methods, given multiple perspectives on tool usage, they are presented as providing a compelling case for validity and generality. However, they would require substantial investment in longitudinal ethnographic studies of large groups which may not be forthcoming.

In the context of evaluation, Pickering et al. [48] recently proposed a step-wedge cluster randomised trial. This was to test the impact of their system, AWARE (Ambient Warning and Response Evaluation), on information management and workflow on a live clinical intensive care unit setting. Such trials are not commonly conducted, but can give real measures of efficiency of data utilisation and may be a good method of evaluation. Their outcome was connected with time spent in data gathering with and without the system and measures were gathered by direct observation and survey.

4 Visualisation of Patient-Centric Pathways

The development of patient-centric pathways and related visualisation tools was first conceptualised as a way to plot and study biomarker trends over time for individual patients with a specific condition. This was carried out in a case study on prostate cancer, where the Prostate Specific Antigen (PSA) was the biomarker test used. The PSA is typically used to measure activity of the cells in the prostate, both benign or malignant, and guidelines for the management and screening of prostate cancer suggest that the PSA test can be read at certain time points to help understand disease progression. As a result, a typical patient

will have several PSA readings during their journey through care and in their pathways.

4.1 Pathways

A pathway is comprised of activities each containing the patient identifier, the event code from a pre-defined dictionary of codes, the time when the activity occurred (in days, zeroed at diagnosis date) and the value pertaining to that specific activity. For example, activity A_4 at time 105 (days after diagnosis) describing the surgical removal of the prostate (event code S) for patient id 8 would be described as $A_4 = (8, 105, S, \text{"M61.1"})$. In this example, the value pertaining to surgical activity code S is the procedure code for the type of surgical operation. We used the OPCS 4.5 Classification of Interventions and Procedures coding and, in this case, code M61.1 refers to a total excision of prostate and its capsule. The activity in this example would, in turn, be part of a pathway, illustrated in Table 1. The pathway data model is defined in more detail in [3].

4.2 Development of a Graph Plotting System

A first support system was developed to plot the biomarker trends based on the pathways data model [3]. This allowed the computation of charts showing the complete PSA trend for each patient in the dataset. The resulting charts were then divided by treatment type and this provided interesting results and posed additional clinical questions. Analysis of the charts, working together with the clinical team, was critical to determine further system requirements and future developments, including a novel graphical representation of pathways data, described later. The data model can be revisited and data elements can be added or removed, making this approach reproducible in other clinical domains and extensible to different levels of granularity.

The inspection of PSA trend plots made clear that these should contain additional information in order to explain, for example, why the biomarker values dropped from abnormal to normal levels at particular points in time. For example, the most significant drops in PSA should be associated with a particular radical treatment. This led to the development of a more sophisticated visualisation system, capable of interpreting the pathways and transforming them into meaningful yet concise graphical representations. The purpose of such visualizations is to summarise complex clinical information over large periods of time into a single graph.

A graph generating system was developed together with the pathways engine, and comprised an architecture similar to that of the Model-view-controller [49] (MVC). In this implementation, the architecture, specific for building graphical representations of pathways, encompasses the following elements with specific purposes:

Table 1. Annotated example of a pathway for patient id 8 with 7 activities and 4 distinct data elements (code P - PSA test, D - Diagnosis, G - Histological Gleason Grade and S - Surgery).

Pathway	Activity	Time	Code	Value	Description
$P = \langle$ $A_1 = (8, -51, P, 13.6),$ $A_2 = (8, 0, D, 2),$ $A_3 = (8, 1, G, "4+3"),$ $A_4 = (8, 105, S, "M61.1"),$ $A_5 = (8, 106, G, "3+4"),$ $A_6 = (8, 183, P, 0.05),$ $A_7 = (8, 456, P, 0.05)$ \rangle	A_1	-51	P	13.6	This patient's first activity was a PSA test (values in ng/ml). In this case the reading was abnormal (>4 ng/ml) 51 days before diagnosis.
	A_2	0	D	2	Diagnosis event, value shows tumour staging. In this case stage 2 indicates the tumour is confined to the prostate capsule. At this point, a biopsy was undertaken (poorly recorded in our EMR systems at the time).
	A_3	1	G	4+3	The result of the histological assessment of the Gleason grade, that is, the degree of cell differentiation, in this case a Gleason sum of 7.
	A_4	105	S	M61.1	The patient then underwent surgery, with an OPCS procedure code of M61.1 indicating total excision of prostate and its capsule.
	A_5	106	G	3+4	The revised Gleason grade with a more complete sample taken from the surgical operation was still a Gleason sum of 7 but now predominantly showing more of type 3 than type 4.
	A_6	183	P	0.05	Post-treatment PSA test was carried out showing a value less than 0.1, denoting effective treatment in reducing the amount of PSA produced in this patient.
	A_7	456	P	0.05	Follow-up PSA test reaffirming that the treatment was successful around a year after the treatment was performed.

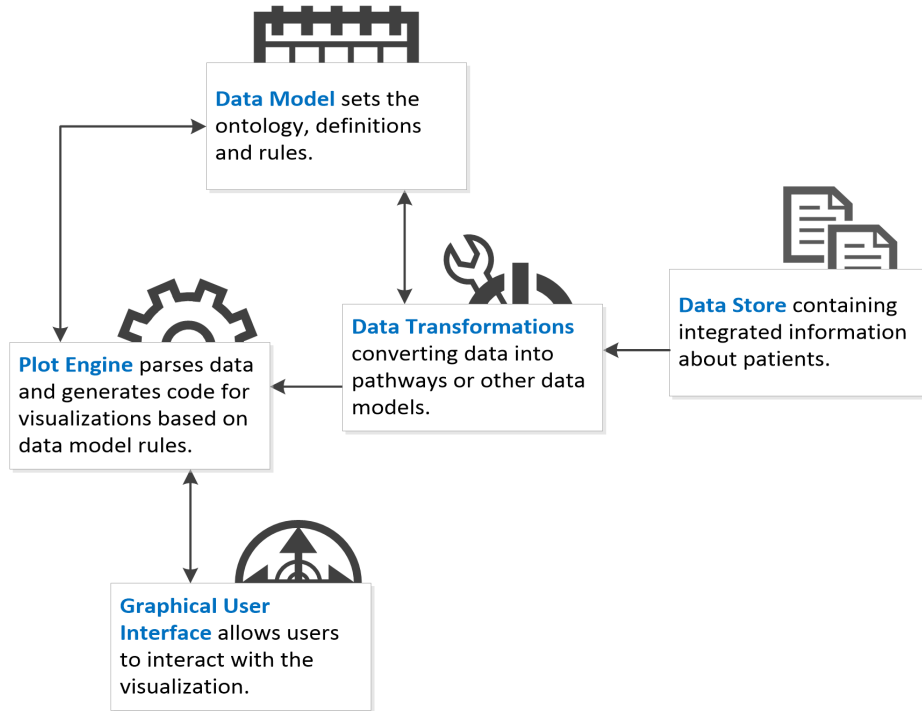


Fig. 1. Architecture of the graph generating system.

- the Data Model, responsible for maintaining the definitions and rules for the interpretation of the pathways data using an extended dictionary that contains information on how events are drawn;
- the Plot Engine, a controller that communicates user or system requests and is responsible for the interaction between the model, the view and the system;
- the Graphical User Interface (containing the view), that receives instructions based on the model and generates a graphical representation of a pathway. This dynamic interface can also allow users to interact with the graphs by communicating information back to the engine.

Figure 1 depicts the architecture of the system. Information available from a Data Store is transformed according to definitions set out by the Data Model and it is then fed to the Plot Engine. In turn, the engine utilises rules on how to draw the graph that is ultimately sent to the Graphical User Interface.

4.3 Graphical Representation

Figure 2 shows the layout of a graph, or pathway plot, and the areas of the graph where information is displayed. The y-axis represents the biomarker values (in

this case, PSA) and the x-axis represents time, in days, zeroed at diagnosis date. The biomarker readings are plotted in the center and events (such as treatments or death) are marked with a vertical line (*Line*).

Treatments and other events can be colour-coded and, above the plot, the corresponding pathway code (e.g. S for Surgery) is shown in the *Line headings* area. The *footer area* displays additional information pertaining to events (such as Gleason grades, i.e. the level of cell differentiation seen in the biopsy, or patient age at diagnosis) and the *right column area* on the right of the plot displays additional information on the patient that is not time-dependent, such as deprivation score, additional diagnoses or alerts.

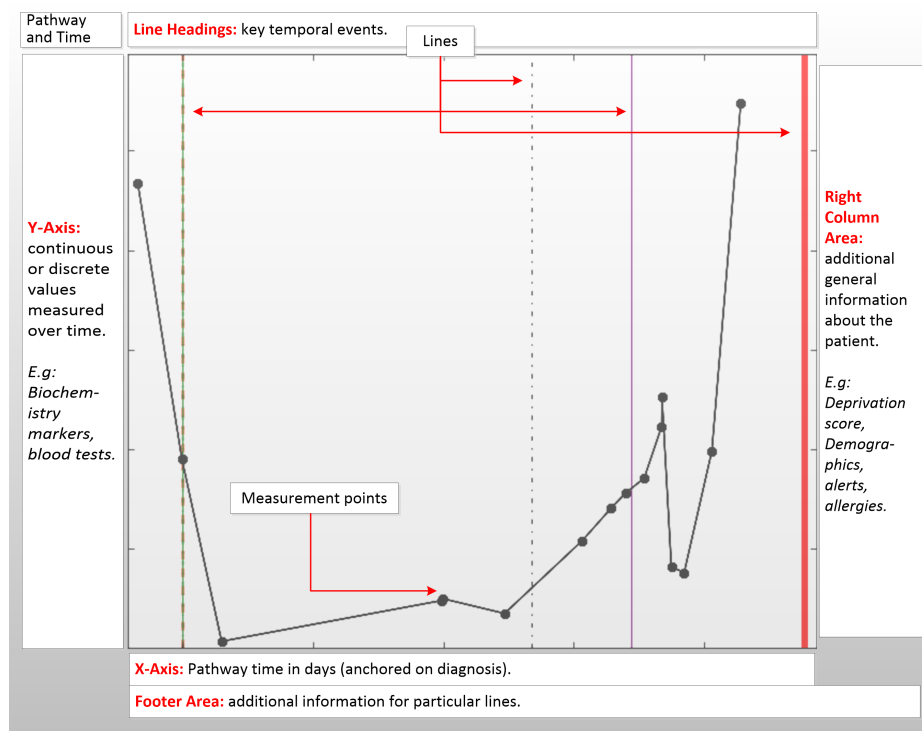


Fig. 2. The schematic layout of a pathway plot.

The graph generating system includes additional interaction capabilities and analysis tools. Rather than relying on static graphical representations of the pathways, the MVC architecture embedded within the system, produces real-time plots of the pathways, as they are read from the database. Dynamic interactions were also introduced enabling users to zoom in, re-scale and navigate the pathway plot. This is particularly important as the scales of the plots may render some drawn objects too close to each other. A mechanism for graphical conflict

resolution (i.e. avoiding overlapping elements) was also introduced. Examples of pathway plots produced by this system are given in section 6.

5 Visualization of Aggregated Pathways

We now explore how to aggregate pathways in a visualisation. The pathways data model enables the production of succinct sequences of activity codes. Truncating the sequence strings (i.e. collapsing sequentially repeating elements into one) enables the aggregation of pathways with similar sequential activities. We developed a web-based software, called ExploraTree, to produce and display an interactive tree of the full cohort of prostate cancer patients based on the available data elements. The technologies used include HTML, CSS, JSON, JavaScript and the InfoVis toolkit. The pathways engine was used to produce the correct data format for a tree representation using JSON and the JavaScript InfoVis toolkit.

In order to accurately aggregate patients with similar sequences of activities, new data elements were introduced in the data dictionary. In the core data dictionary, a patient's death was encoded by only one data element (code Z). In the new encoding, patients who died of prostate cancer were kept with code Z while those who died of other causes were identified with code Y and those who survived, with code X. This ensures that all patients have a terminal element indicating whether they are alive at the end of their follow-up period. Because in this cohort not all patients are followed-up the same amount of time, all terminal elements (X,Y,Z) were given additional child nodes that represent the amount of time the patients were followed-up in years (1 to 5 and '+' for over 5 years). The aggregated pathways tree is illustrated in Figure 3.

Figure 3 shows the cohort tree and highlighted sequence $\langle P, D, H, P, X \rangle$, that is, patients who started their pathway with one or more PSA tests (code P, $n=1596$), followed by a diagnosis of cancer (code D, $n=1502$), hormone therapy as first treatment (code H, $n=747$), other PSA test(s) ($n=557$) and finally were last seen alive in this cohort (code X). 90% of patients with the highlighted pathway ($n=266$) were followed-up 3 or more years and one patient was followed-up less than one year.

This aggregation also allows comparing patients that followed similar pathways but who died of prostate cancer ($\langle P, D, H, P, Z \rangle$). In the case of patients with a sequence prefix $\langle P, D, H, P \rangle$, 9% ($n=48$) died of prostate cancer (code Z), 13% died of other causes (code Y), 48% survived, and the remaining patients continued with other activities (H - Hormone Therapy, W - Active Surveillance, R - Radiotherapy, S - Surgery).

Visualising the cohort in this manner is important as it enables the selection of subsets of data for specific clinical studies as well as an inspection of the sequential routes that patients take through care. The sequence highlighted in Figure 3 corresponds to the most common route (with most support on each node sequentially).

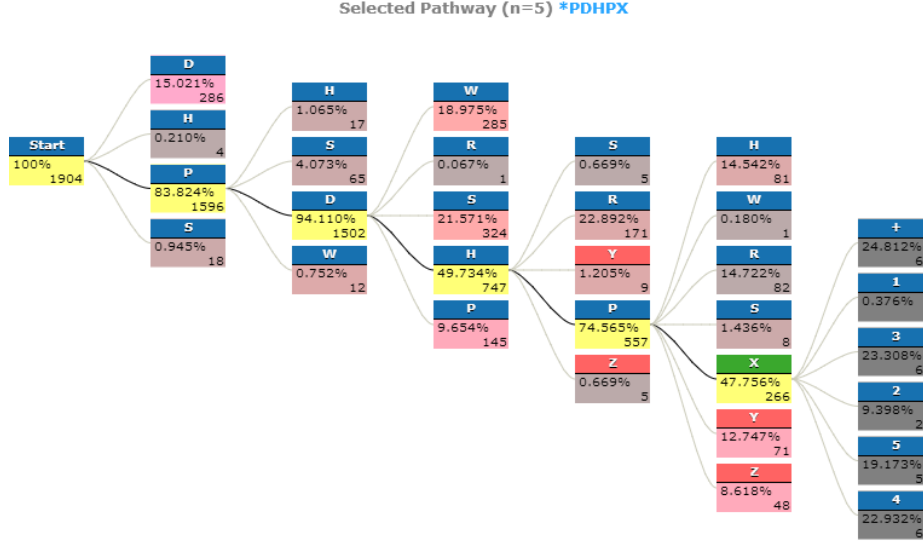


Fig. 3. CaP VIS ExploraTree software displaying a selected pathway (patients with the same sequential activities). The selected pathway nodes are highlighted and terminal nodes are marked as red for patients that died and green for patients that were last seen alive in this cohort.

It is possible to add more meaning to the visualisation and the pathways by introducing additional data elements and remodeling the data dictionary. For example, instead of using a single code for diagnosis it is possible to have a breakdown of the tumour staging or Gleason grade at diagnosis so as to group similar sequences with this information instead. However, due to the small size of this cohort, increasing granularity in the pathways dictionary would result in fewer patients in each node. For this reason no additional changes were made to the pathways dictionary used for the ExploraTree, but our approach is flexible enough to allow such modifications.

6 Application Areas

This section lists four broad areas where visualisation tools have been applied and are expected to be most useful. Pathway plots illustrating relevant examples are given for each of the areas.

6.1 Decision Support and EMR Enhancement

Recommendations for further research in clinical decision support and expert systems [50] suggest that software that integrates complex data and generates graphical representations is needed to support the analysis and understanding

of the data. Visualisations could also be used to enhance EMR systems as these do not typically provide visually meaningful summaries of patient-centric data.

The pathways software was developed so that additional clinical information, such as histopathology text reports, descriptive statistics, and graphical representation could all be available in one place. This created an environment that enables evidence based medicine, supports decision making. Clinicians are able to retrieve similar cases by searching the desired pathway sequences and visually inspect them, thereby gaining insights to support their decisions. In addition, other information derived from domain knowledge such as PSA kinetics (how fast PSA readings are doubling in time and rate of increase, both predictive of outcome) can be shown in the developed system before or after diagnosis and treatment. The flexible pathways data model has also enabled other aspects to be incorporated. For example, rules can be applied to measure adherence to guidelines.

Figure 4 shows four pathway plots for the same patient, a 69 year old diagnosed with tumour stage 3 prostate cancer and a Gleason sum of 9. Plot A shows the original plot where the PSA is seen to have dropped after the patient underwent hormone therapy (code H). The thick red line at the end of the pathway denotes when the patient died. When producing this pathway's plots, the dictionary was extended so that the treatments retrieved from the local cancer registry (and additional source of validation data) appear with a suffix "1" in the vertical lines' headings (code H1). In this case, regarding the date when the patient first commenced hormone therapy, a time discrepancy of 51 days was seen between the two data sources, where the hospital recorded the later date. Hence this serves to inform on data quality issues (further discussed in the next section). The discrepancy in dates in this case did not introduce uncertainty as the effect of the treatment is seen in the subsequent PSA readings.

The pathway plot in Figure 4 then shows a PSA relapse in the last two readings. Shortly after the last PSA reading, the patient died of a pulmonary embolism (ICD I26) and prostate cancer (ICD C61) as a secondary condition leading to death. Shortly before death the patient was diagnosed with a secondary and unspecified malignant neoplasm of inguinal and lower limb nodes (ICD C77.4). This was revealed by the additional data collected on hospital episodes and is presented in the visualisation.

Figure 4 Plot B shows an additional element of the pathway, a blood test, Alkaline Phosphatase (ALP) and its normal range in the shaded area. When a patient's advanced cancer metastasises to the bones, ALP can be increased due to active bone formation. Indeed studies have shown that prostate cancer patients with a serum ALP reading of more than twice the normal upper limit had a significantly lower survival rate than their respective counterparts [51]. This is observed in this pathway, although, an increased ALP could be due to other reasons such as an obstructed bile duct or liver disease.

Lastly, plots C and D supplement the pathway with another blood test, Creatinine. Creatinine has been reportedly associated with more advanced disease and decreased survival [52]. However, any condition that impairs the function

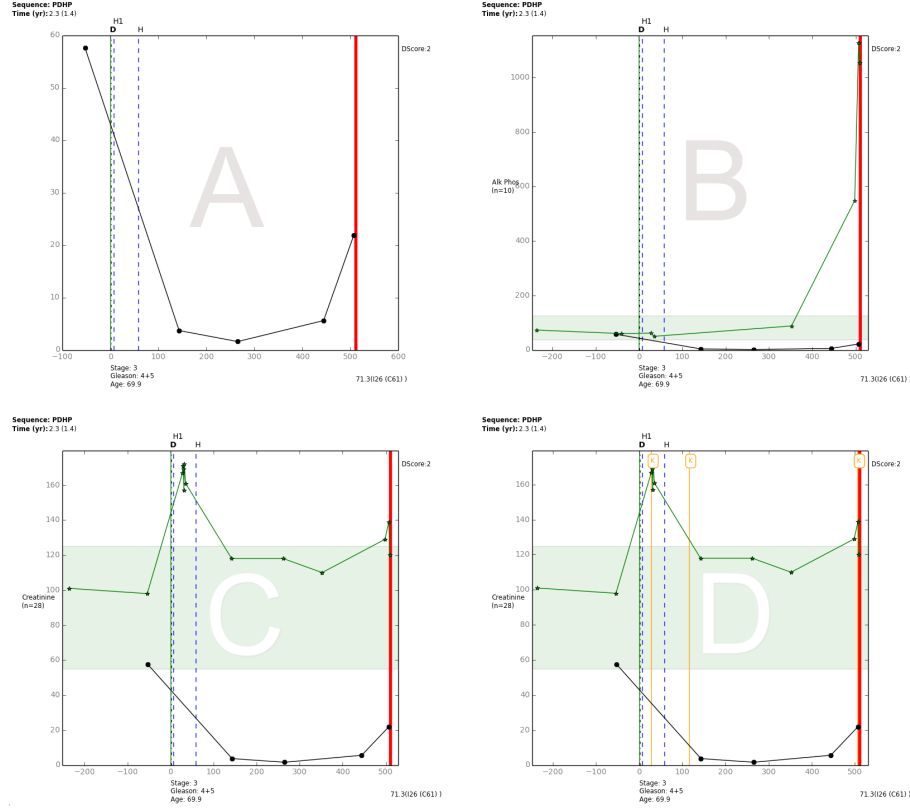


Fig. 4. Four pathway plots of the same patient (175) with sequence $\langle P, D, H, P \rangle$. Plot A shows the original plot with the PSA trend alone. Plot B shows the same information as plot A with additional Alkaline Phosphatase readings and their normal range (shaded area). Plot C shows Creatinine readings and Plot D shows the same information and hospital events (code K).

of the kidneys is likely to raise the creatinine levels in the blood and act as a confounding factor. In plot C, a flare in the values of Creatinine readings was observed within the first 3 months. By introducing additional data elements from the hospital episode statistics in plot D, a hospital episode (marked with pathway code K) was found with an associated primary diagnosis of acute kidney failure. Additional detail on episodes is obtainable by interacting with the visualisation. Although a kidney stone was not coded in this (or any) episode for this patient, a catheterisation of the bladder was performed during the same hospital visit, and an inspection of the patient notes confirmed a kidney stone was the cause of the acute kidney failure. The second hospital episode in this pathway, also marked with code K, was for the removal of the catheter, and the last hospital episode included a diagnosis of a secondary and unspecified ma-

lignant neoplasm of inguinal and lower limb nodes and a pulmonary embolism, caused by the first. This level of information that can be added to the pathway would also allow, for example in other cases, to evaluate renal impairment and prostate cancer. Indeed, in this respect, it has been reported that renal impairment in men undergoing prostatectomy represents substantial and unrecognised morbidity [53].

The introduction of additional detail helped to explain the Creatinine flare for this patient and provided interesting insights that would otherwise not be easily explored. The pathway plots provided sufficient information for the interpretation of the pathway yet highlighted potential issues with the quality of the data. Indeed discrepancies in treatment dates across data sources may introduce additional challenges. As such, it is important to be able to differentiate between pathways that have sufficient information and provide an accurate representation of the patient’s history and those that do not. The evaluation of the completeness and utility of the generated pathways for investigating biomarker trends is explored in more detail in the next section.

6.2 Data Quality

Methods for the evaluation of data quality dimensions are lacking [54] and visualisation tools can play an important role in quality assurance. Since the development of the pathways framework, one of the first and foremost concerns pertained to the quality of the data being visualised. For the first time since EMR systems were introduced in our hospital, it was possible to visualise integrated data and observe inconsistencies in the ways in which information had been recorded over time. By expanding the data dictionary to include additional information from an external data source, the regional Cancer Registry, it was possible to identify incongruent data across sources.

Figure 5 shows a pathway plot of a patient with Gleason grade 7 prostate cancer who underwent a radical prostatectomy. Information from the Cancer Registry was obtained to validate treatment data and this is included with code S1. In this case, the dates and details of the procedure are in agreement and this patient could easily pass for having a complete record. When plotting the pathway, however, a visual inspection highlighted a significant drop in the PSA values for which there is no clear justification based on the information available. It is unlikely that the PSA values dropped below the 4 ng/ml normal threshold without an intervention. This means that either the treatment date is incorrect in both sources or there is missing information as the patient is likely to have received treatment from another provider while the blood tests continued to be performed by the same laboratory. In this case the plausibility and concordance data quality dimensions were assessed with this visualisation.

Other data quality examples include mismatch of treatment dates (as seen earlier in Figure 4) and missing or implausible information. Based on the pathways framework, rules can be devised to inspect individual pathways and determine how complete they might be. For example, in previous work [3] rules pertaining to the availability, positioning and substantiation of the drops in PSA

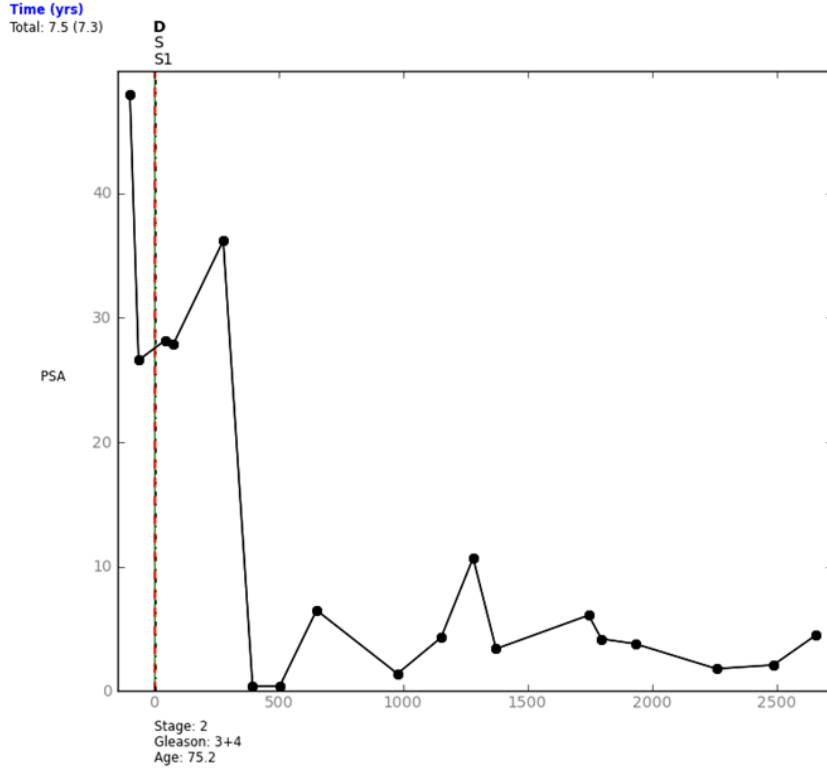


Fig. 5. A Pathway plot for a patient diagnosed with Gleason grade 7 prostate cancer who underwent a radical prostatectomy (code S).

were proposed to determine which pathways would be eligible for further clinical research.

6.3 Cohort Selection, Analysis and Research

Two of the preliminary interests in developing graphical representations of pathways were to compare the shapes of the biomarker curves and also to be able to aggregate patients with similar features. Having pathways expressed as sequences of activity codes has helped to develop the ExploraTree tool, seen in Figure 3. Depending on how the data points and outcomes are modelled, the trees produced will have varying degrees of granularity and clinical interest. In the example shown earlier, ExploraTree is aggregating patients with similar data points appearing sequentially in time. However, codes for PSA tests (P) could be further broken down into abnormal (say, A) and normal (N) PSA values and this would create more clinically meaningful groups. The ExploraTree software can then help to select relevant cohorts for research, to determine if there are

enough members in a particular group of interest and to facilitate recruitment for prospective studies.

Pathway plots allow more detailed and complex information to be presented in a single graphical representation. This enables researchers to observe several data points together and to study new outcomes. For example, Figure 6 plots Haemoglobin in addition to the PSA and shows normal perioperative bleeding when the patient underwent surgery. This information is not usually examined together yet it enables the assessment of the effect that surgical procedures have on patients and also, the length of time it takes for them to recover after surgery. The latter is an interesting current research question that arises from the visual inspection of the pathways. It is also possible to determine and study different outcomes such as hormone escaped, development of metastases or biochemical recurrence after treatment. Research on services and adherence to guidelines is also possible using the pathway framework [3]. Integration of clinical EMR data with “omics” data is also a topic that should deserve attention in future developments. Pathways with this additional information can be more valuable for precision medicine and their visualisations should also help take knowledge of clinical practice out of the hospitals and bring it to biologists, geneticists and other scientists.

6.4 Knowledge Discovery Support

Visualisation tools are often overlooked when working on knowledge discovery problems in healthcare. One of the most common barriers in machine learning in healthcare is that the models and results produced are not intelligible and work in this area is becoming more topical [55]. Decision trees continue to be the gold standard of intelligible models and more work is needed to create visualisation tools that describe complex models.

Data and process mining techniques are often suggested for the analysis of workflows and pathways, however, most of these techniques have been found unsuitable when applied to heterogeneous routine clinical data. The evaluation of the quality of event logs in process mining relies on trustworthiness (recorded events actually happened), completeness and well defined semantics [56]. These can be achieved by selecting pathways with required data points using the pathways framework. The visualisation system allows for the close inspection and contextualisation of pathways, illustrating particular paths with similar features. It has been reported that a combination of visual analytics with automated process mining techniques would make possible the extraction of more novel insights from event data [56] and further work in this area is needed.

The pathways framework through its graphical representations could also be an interesting way of representing a model, whereby an ideal pathway would be presented and then compared to actual pathways and deviation could be measured, although further work in this area is required. Additional analysis of the shape of the curves represented (for example, clustering of biomarker trends) is also possible using this framework and some work has already been done in this area using fusion methods [57].

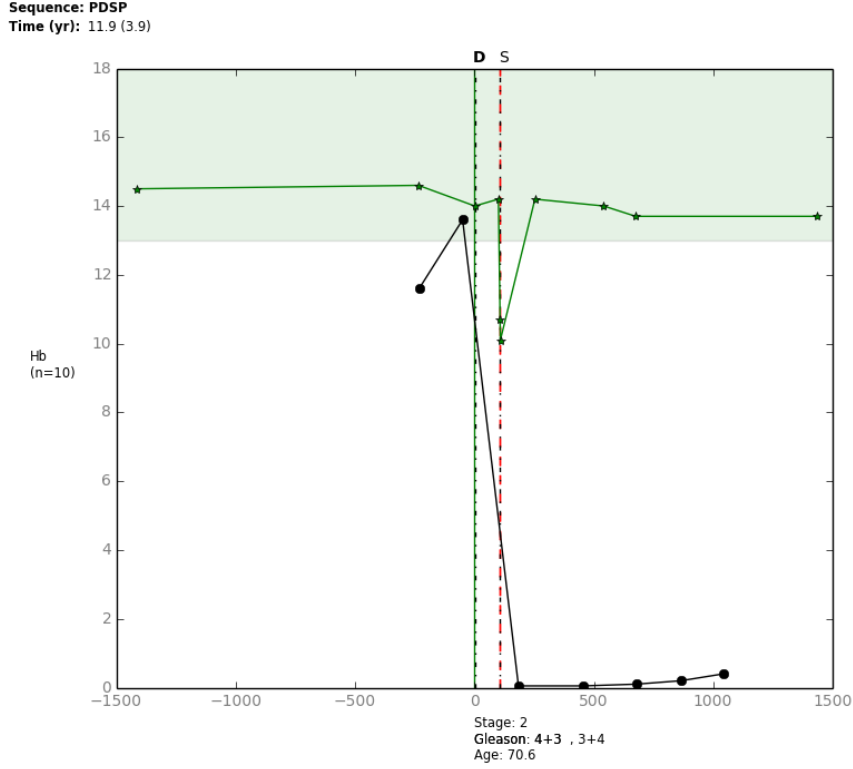


Fig. 6. A Pathway plot showing the effect of a prostatectomy in the Haemoglobin and PSA readings. The green shaded area depicts the normal range for Haemoglobin.

7 Open Problems

Some of the main problems relating to the improvement of health and healthcare with interactive visualisation methods are reviewed by Shneiderman et al. [6], Aigner et al. [58], Caban and Gotz [20], and West et al. [19]. Some of these challenges arise because healthcare must become more “predictive, preemptive, personalised and participative” [6]. Although the efforts described in section 3 and our own efforts are directed to some of these challenges, most systems described do not provide completely satisfactory responses. The open problems summarised from the papers above and from the work presented here include:

- An enduring problem in visualising clinical data is the scale and complexity of the data. Data is not only vast in terms of the number of records but it also includes several different data types (e.g. numeric, categorical, text, images), semantic structures inherent of time data such as cycles and re-occurrences and intertwining conditions and treatment processes. Visual techniques must analyse data in the context of this complexity and summarise it in order to

assist busy clinicians with getting timely information in the right format. This requires tools that enable the user to see the overall perspective with powerful yet simple visualisations and then look for anomalies and drill for details of predictable risks early.

- The systems must be capable of scaling up to cohort analysis. Visualising one patient’s trajectory can enable monitoring of treatment process for that particular patient. However, it is often necessary to scale the analysis to a cohort of patients as clinicians can then compare responses of diverse patients and assess effectiveness of therapy in the larger scale.
- Context and domain knowledge is very important in clinical decision making so systems must be able to efficiently represent domain knowledge and reason with it to make temporal abstractions, to look at conditions in the context of many clinical parameters such as co-morbidities, medication and history. It may also be desirable to compare cohorts across clinicians, time periods and geographical locations.
- It is increasingly necessary to provide systems that can facilitate multi-disciplinary decision making. Such teams may involve nurses, social workers, physicians and patients. Hence the presentation of knowledge, flexible querying and analysis should accommodate the demands of multiple users with different perspectives and needs. Visualisation tools should play an important role in delivering and interacting with patient data.
- It is often necessary to understand similarity in the context of heterogeneous data but this is not a well developed area of research. Data mining tasks such as classification, clustering, association rules and deviation detection need to be developed to work with heterogeneous temporal data and to produce intelligible results and meaningful visualisations.
- Data that is routinely collected is plagued by missing values, erroneous values and inaccuracies. Systems that analyse such data must be well equipped to deal with uncertainty. However, uncertainty is a well known open problem in computing. Issues of data quality take their own dimension in a time oriented scenario and can require specific treatment [59]. It is necessary to pre-process the data to uncover data quality issues and exclude dubious data from further analysis. It is also important to quantify data quality dimensions by producing standard measures that can be presented (visually) alongside the data. In addition, presentation of uncertainty in a meaningful way, for example in the context of risk, is still an open research area.
- Currently, according to Kopanitsa et al. [60], there is a gap in transforming knowledge from domain model to interface model. Hence there is a need to turn hard-coded user interfaces into generic methods by a process of standardisation. Standardisation exists for data storage and exchange and they provide a good basis for further efforts. This may also make data more accessible to patients, which may be an important consideration for personalised and participative medicine.
- The design of better interfaces was highlighted as a challenge early on [61] and continues to be an open issue. In particular application of cognitive engineering methods [62] may be beneficial for informing design and for

uncovering information needs in clinical systems. There is a requirement for analysing and understanding the process of visual interaction, for example by using logs. Interaction with the visualisation tools is key and must cater for different types of users with different priorities as already discussed.

8 Conclusion and Future Outlook

A picture can arguably be worth a thousand words and in the case of the pathways, a pathway plot is worth, on average, 188 activities using our prostate cancer cohort. For immediate decision-making by clinicians at the point of care, information should be brief and easily interpreted [63] and visualisation tools, if well designed, have a great potential to become part of clinical practice by summarising complex activities in one graphical representation. However, optimal visualisation of clinical data is complex and several open problems remain.

In this paper, clinical pathways were used to demonstrate the potential of visualising routinely collected data using a case study on prostate cancer. The underlying data model enables the summarisation and extension of pathways as well as the aggregation of similar sequences. It is also possible to capture and plot pathways with concurrent elements and to develop algorithms to further explore the data and investigate quality issues. Furthermore, the pathways framework has facilitated interpretation, communication and debate between experts. More work is now needed to assess similar tools in other settings and domains. In this paper, four key areas that hold promise in the future of visualisation in healthcare were identified: decision support and EMR enhancement; data quality; cohort selection, analysis and research; and knowledge discovery. Further work in each of these areas will bring clinical practice closer to the best available evidence and improve the quality and utility of the *big data* that is available in EMR systems.

References

1. Mannu, G.S., Kyu, M.M., Bettencourt-Silva, J.H., Loke, Y.K., Clark, A.B., Metcalf, A.K., Potter, J.F., Myint, P.K.: Age but not abcd2 score predicts any level of carotid stenosis in either symptomatic or asymptomatic side in transient ischaemic attack. *International Journal of Clinical Practice* **69**(9) (2015) 948–956
2. Bettencourt-Silva, J., De La Iglesia, B., Donell, S., Rayward-Smith, V.: On creating a patient-centric database from multiple hospital information systems. *Methods of Information in Medicine* **51**(3) (2012) 210–220
3. Bettencourt-Silva, J.H., Clark, J., Cooper, C.S., Mills, R., Rayward-Smith, V.J., de la Iglesia, B.: Building data-driven pathways from routinely collected hospital data: A case study on prostate cancer. *JMIR Medical Informatics* **3**(3) (Jul-Sep 2015) e26
4. Huang, Z., Lu, X., Duan, H.: On mining clinical pathway patterns from medical behaviors. *Artificial Intelligence in Medicine* **56**(1) (2012) 35–50
5. Zhang, Y., Padman, R., Patel, N.: Paving the cowpath: Learning and visualizing clinical pathways from electronic health record data. *Journal of Biomedical Informatics* **58** (2015) 186 – 197

6. Shneiderman, B., Plaisant, C., Hesse, B.W.: Improving healthcare with interactive visualization. *Computer* **46**(5) (2013) 58–66
7. Potamias, G.: State of the art on systems for data analysis, information retrieval and decision support. INFOBIOMED project (Deliverable D1) (2006)
8. HiMSS: Healthcare information and management systems society(HiMSS). <http://www.himss.org/clinical-informatics/medical-informatics> Accessed: 2015-12-30.
9. Dell Ltd: Data mining techniques. <http://documents.software.dell.com/statistics/textbook/data-mining-techniques> Accessed: 2015-12-30.
10. Perer, A., Wang, F.: Frequence: Interactive mining and visualization of temporal frequent event sequences. In: Proceedings of the 19th International Conference on Intelligent User Interfaces. IUI '14, New York, NY, USA, ACM (2014) 153–162
11. Thomas, J.J., Cook, K.A.: Illuminating the Path: The Research and Development Agenda for Visual Analytics. National Visualization and Analytics Ctr (2005)
12. Batini, C., Scannapieco, M.: Data Quality: Concepts, Methodologies and Techniques. Springer, Berlin, Heidelberg, New York (2006)
13. Hunter, B., Segrott, J.: Re-mapping client journeys and professional identities: A review of the literature on clinical pathways. *International Journal of Nursing Studies* **45**(4) (2008) 608 – 625
14. Shahar, Y., Goren-Bar, D., Boaz, D., Tahan, G.: Distributed, intelligent, interactive visualization and exploration of time-oriented clinical data and their abstractions. *Artif. Intell. Med.* **38**(2) (October 2006) 115–135
15. Field, M.J., Kathleen N. Lohr, E.: Guidelines for Clinical Practice. Institute of Medicine, National Academy Press, Washington, D.C. (1992) An optional note.
16. Turkay, C., Jeanquartier, F., Holzinger, A., Hauser, H.: On Computationally-Enhanced Visual Analysis of Heterogeneous Data and Its Application in Biomedical Informatics. In: Interactive Knowledge Discovery and Data Mining in Biomedical Informatics: State-of-the-Art and Future Challenges. Springer Berlin Heidelberg, Berlin, Heidelberg (2014) 117–140
17. Rind, A., Wang, T.D., Aigner, W., Miksch, S., Wongsuphasawat, K., Plaisant, C., Shneiderman, B.: Interactive information visualization to explore and query electronic health records. *Foundations and Trends in Human-Computer Interaction* **5**(3) (2011) 207–298
18. Roque, F.S., Slaughter, L., Tkatšenko, A.: A comparison of several key information visualization systems for secondary use of electronic health record content. In: Proceedings of the NAACL HLT 2010 Second Louhi Workshop on Text and Data Mining of Health Documents. Louhi '10, Stroudsburg, PA, USA, Association for Computational Linguistics (2010) 76–83
19. West, V.L., Borland, D., Hammond, W.E.: Innovative information visualization of electronic health record data: a systematic review. *Journal of the American Medical Informatics Association* (2014)
20. Caban, J.J., Gotz, D.: Visual analytics in healthcare – opportunities and research challenges. *Journal of the American Medical Informatics Association* **22**(2) (2015) 260–262
21. Lesselroth, B.J., Pieczkiewicz, D.S.: Data visualization strategies for the electronic health record. In Berhardt, L.V., ed.: *Advances in Medicine and Biology*. Volume 16. Nova Science Publisher Inc. (2012) 107–140
22. Aigner, W., Miksch, S., Schuman, H., Tominski, C.: Visualization of Time-Oriented Data. 1st edn. Human-Computer Interaction. Springer Verlag (2011)

23. Plaisant, C., Milash, B., Rose, A., Widoff, S., Shneiderman, B.: Lifelines: Visualizing personal histories. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '96, New York, NY, USA, ACM (1996) 221–227
24. Plaisant, C., Mushlin, R., Snyder, A., Li, J., Heller, D., Shneiderman, B.: Lifelines: using visualization to enhance navigation and analysis of patient records. *Proceedings of the AMIA Symposium* (1998) 76–80
25. Wang, T.D., Plaisant, C., Quinn, A.J., Stanchak, R., Murphy, S., Shneiderman, B.: Aligning temporal data by sentinel events: Discovering patterns in electronic health records. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '08, New York, NY, USA, ACM (2008) 457–466
26. Wongsuphasawat, K., Guerra Gómez, J.A., Plaisant, C., Wang, T.D., Taieb-Maimon, M., Shneiderman, B.: Lifeflow: Visualizing an overview of event sequences. In: *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. CHI '11, New York, NY, USA, ACM (2011) 1747–1756
27. Shahar, Y., Cheng, C.: Intelligent visualization and exploration of time-oriented clinical data. In: *Systems Sciences, 1999. HICSS-32. Proceedings of the 32nd Annual Hawaii International Conference on*. Volume Track4. (Jan 1999) 12 pp.–
28. Klimov, D., Shahar, Y., Taieb-Maimon, M.: Intelligent visualization and exploration of time-oriented data of multiple patients. *Artificial Intelligence in Medicine* **49**(1) (2010) 11 – 31
29. Duke, J.D., Bolchini, D.: A successful model and visual design for creating context-aware drug-drug interaction alerts. *AMIA Annu Symp Proc* **2011** (2011) 339–348
30. Huang, C.W., Lu, R., Iqbal, U., Lin, S.H., Nguyen, P.A.A., Yang, H.C., Wang, C.F., Li, J., Ma, K.L., Li, Y.C.J., Jian, W.S.: A richly interactive exploratory data analysis and visualization tool using electronic medical records. *BMC Medical Informatics and Decision Making* **15**(1) (2015) 1–14
31. Riehmann, P., Hanfler, M., Froehlich, B.: Interactive sankey diagrams. In: *Information Visualization, 2005. INFOVIS 2005. IEEE Symposium on*. (Oct 2005) 233–240
32. Wong, B.L.W., Xu, K., Holzinger, A.: Interactive Visualization for Information Analysis in Medical Diagnosis. In: *Information Quality in e-Health: 7th Conference of the Workgroup Human-Computer Interaction and Usability Engineering of the Austrian Computer Society, USAB 2011, Graz, Austria, November 25-26, 2011. Proceedings*. Springer Berlin Heidelberg, Berlin, Heidelberg (2011) 109–120
33. Aigner, W., Miksch, S.: Carevis: Integrated visualization of computerized protocols and temporal patient data. *Artificial Intelligence in Medicine* **37**(3) (2006) 203 – 218 *Knowledge-Based Data Analysis in Medicine*.
34. Aigner, W., Miksch, S.: Supporting protocol-based care in medicine via multiple coordinated views. In: *Coordinated and Multiple Views in Exploratory Visualization, 2004. Proceedings. Second International Conference on*. (July 2004) 118–129
35. Bodesinsky, P., Federico, P., Miksch, S.: Visual analysis of compliance with clinical guidelines. In: *Proceedings of the 13th International Conference on Knowledge Management and Knowledge Technologies. i-Know '13, New York, NY, USA, ACM* (2013) 12:1–12:8
36. Krause, J., Perer, A., Stavropoulos, H.: Supporting iterative cohort construction with visual temporal queries. *IEEE Trans Vis Comput Graph* **22**(1) (Jan 2016) 91–100
37. Gotz, D., Stavropoulos, H.: Decisionflow: Visual analytics for high-dimensional temporal event sequence data. *Visualization and Computer Graphics, IEEE Transactions on* **20**(12) (Dec 2014) 1783–1792

38. Kamsu-Foguem, B., Tchuente-Foguem, G., Allart, L., Zennir, Y., Vilhelm, C., Mehdaoui, H., Zitouni, D., Hubert, H., Lemdani, M., Ravaux, P.: User-centered visual analysis using a hybrid reasoning architecture for intensive care units. *Decision Support Systems* **54**(1) (2012) 496 – 509
39. Stolper, C., Perer, A., Gotz, D.: Progressive visual analytics: User-driven visual exploration of in-progress analytics. *IEEE Transactions on Visualization and Computer Graphics* **20**(12) (12 2014) 1653–1662
40. Sturm, W., Schreck, T., Holzinger, A., Ullrich, T.: Discovering Medical Knowledge Using Visual Analytics. In Buhler, K., Linsen, L., John, N.W., eds.: *Eurographics Workshop on Visual Computing for Biology and Medicine*, The Eurographics Association (2015)
41. Otasek, D., Pastrello, C., Holzinger, A., Jurisica, I.: Visual Data Mining: Effective Exploration of the Biological Universe. In: *Interactive Knowledge Discovery and Data Mining in Biomedical Informatics: State-of-the-Art and Future Challenges*. Springer Berlin Heidelberg, Berlin, Heidelberg (2014) 19–33
42. Lavrac, N., Bohanec, M., Pur, A., Cestnik, B., Debeljak, M., Kobler, A.: Data mining and visualization for decision support and modeling of public health-care resources. *Journal of Biomedical Informatics* **40**(4) (2007) 438 – 447 *Public Health Informatics*.
43. Perer, A., Wang, F., Hu, J.: Mining and exploring care pathways from electronic medical records with visual analytics. *Journal of Biomedical Informatics* **56** (2015) 369 – 378
44. Tate, A., Beloff, N., Al-Radwan, B., Wickson, J., Puri, S., Williams, T., van Staa, T., Bleach, A.: Exploiting the potential of large databases of electronic health records for research using rapid search algorithms and an intuitive query interface. *Journal of the American Medical Informatics Association* **21**(2) (2014) 292–298
45. Jeanquartier, F., Jean-Quartier, C., Holzinger, A.: Integrated web visualizations for protein-protein interaction databases. *BMC Bioinformatics* **16**(1) (2015) 1–16
46. Müller, H., Reihs, R., Zatloukal, K., Holzinger, A.: Analysis of biomedical data with multilevel glyphs. *BMC Bioinformatics* **15**(6) (2014) 1–12
47. Shneiderman, B., Plaisant, C.: Strategies for evaluating information visualization tools: Multi-dimensional in-depth long-term case studies. In: *Proceedings of the 2006 AVI Workshop on BEyond Time and Errors: Novel Evaluation Methods for Information Visualization*. BELIV '06, New York, NY, USA, ACM (2006) 1–7
48. Pickering, B.W., Dong, Y., Ahmed, A., Giri, J., Kilickaya, O., Gupta, A., Gajic, O., Herasevich, V.: The implementation of clinician designed, human-centered electronic medical record viewer in the intensive care unit: A pilot step-wedge cluster randomized trial. *International Journal of Medical Informatics* **84**(5) (2015) 299 – 307
49. Buschmann, F., Henney, K., Schmidt, D.: *Pattern-oriented Software Architecture: On Patterns and Pattern Language*. Volume 5. John Wiley & Sons (2007)
50. Berner, E.S.: *Clinical Decision Support Systems: Theory and Practice*. 2nd edn. Springer Publishing Company, Incorporated (2010)
51. Nakashima, J., Ozu, C., Nishiyama, T., Oya, M., Ohigashi, T., Asakura, H., Tachibana, M., Murai, M.: Prognostic value of alkaline phosphatase flare in patients with metastatic prostate cancer treated with endocrine therapy. *Urology* **56**(5) (2000) 843 – 847
52. Weinstein, S.J., Mackrain, K., Stolzenberg-Solomon, R.Z., Selhub, J., Virtamo, J., Albanes, D.: Serum creatinine and prostate cancer risk in a prospective study. *Cancer Epidemiology Biomarkers & Prevention* **18**(10) (2009) 2643–2649

53. HILL, A.M., Philpott, N., Kay, J., Smith, J., Fellows, G., Sacks, S.: Prevalence and outcome of renal impairment at prostatectomy. *British journal of urology* **71**(4) (1993) 464–468
54. Weiskopf, N.G., Weng, C.: Methods and dimensions of electronic health record data quality assessment: enabling reuse for clinical research. *Journal of the American Medical Informatics Association* **20**(1) (2013) 144–151
55. Caruana, R., Lou, Y., Gehrke, J., Koch, P., Sturm, M., Elhadad, N.: Intelligible models for healthcare: Predicting pneumonia risk and hospital 30-day readmission. In: *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. KDD '15*, New York, NY, USA, ACM (2015) 1721–1730
56. Van Der Aalst, W., Adriansyah, A., de Medeiros, A.K.A., Arcieri, F., Baier, T., Blickle, T., Bose, J.C., van den Brand, P., Brandtjen, R., Buijs, J., et al.: Process mining manifesto. In: *Business process management workshops*, Springer (2012) 169–194
57. Mojahed, A., Bettencourt-Silva, J., Wang, W., de la Iglesia, B.: Applying clustering analysis to heterogeneous data using similarity matrix fusion (smf). In Perner, P., ed.: *Machine Learning and Data Mining in Pattern Recognition*. Volume 9166 of *Lecture Notes in Computer Science*. Springer International Publishing (2015) 251–265
58. Aigner, W., Federico, P., Gschwandtner, T., Miksch, S., Rind, A.: Challenges of time-oriented data in visual analytics for healthcare. In Caban, J.J., Gotz, D., eds.: *IEEE VisWeek Workshop on Visual Analytics in Healthcare*, IEEE, IEEE (2012) 4
59. Gschwandtner, T., Gärtner, J., Aigner, W., Miksch, S.: A taxonomy of dirty time-oriented data. In Quirchmayr, G., Basl, J., You, I., Xu, L., Weippl, E., eds.: *Lecture Notes in Computer Science (LNCS 7465): Multidisciplinary Research and Practice for Information Systems (Proceedings of the CD-ARES 2012)*, Springer, Berlin / Heidelberg, Springer, Berlin / Heidelberg (2012) 58 – 72
60. Kopanitsa, G., Hildebrand, C., Stausberg, J., Englmeier, K., et al.: Visualization of medical data based on ehr standards. *Methods Inf Med* **52**(1) (2013) 43–50
61. Tang, P.C., Patel, V.L.: Major issues in user interface design for health professional workstations: summary and recommendations. *International Journal of Bio-Medical Computing* **34**(14) (1994) 139 – 148 *The Health Care Professional Workstation*.
62. Thyvalikakath, T.P., Dziabiak, M.P., Johnson, R., Torres-Urquidy, M.H., Acharya, A., Yabes, J., Schleyer, T.K.: Advancing cognitive engineering methods to support user interface design for electronic health records. *International Journal of Medical Informatics* **83**(4) (2014) 292 – 302
63. Kay, J.D.: Communicating with clinicians. *Annals of clinical biochemistry* **38** (2001) 103