Visualization for Quality Healthcare: Patient Flow Exploration

1st Veronika Domova
Linköping University
Västerås, Sweden
verdo29@liu.se

2nd Shiva Sander-Tavallaey, PhD
ABB Corporate Research
Västerås, Sweden
shiva.sander-tavallaey@se.abb.com

Abstract—The health and medical institutes and hospitals are facing several challenges in the short and long term, i.e. demographic changes, demands on improving quality, limited resources and cost requirements. To cope with these challenges, the health care sector needs to become more efficient, while maintaining and improving the quality of care. Efficient and transparent information flow across the whole system plays a key role even in this context. Being able to analyze the patient flows as the first step will provide the possibility to more efficiently manage medical resources and better serve incoming patients. In this work, we have developed a web-based portal for interactive analysis of patient flow data to assist hospital authorities to improve and optimize the time and quality of the provided services. The developed tool facilitates short and long term optimization of resource allocation by analyzing the past, as well as current, patient flows, identifying bottlenecks and exploring the reasons for the occurred waiting times.

Index Terms—Medical data, Visual analytics, Temporal event visualization, Patient flow, Patient trails

I. INTRODUCTION

In this paper, we present the results from a long-term collaborative medical research project aiming to increase the efficiency of Swedish municipal hospitals. In the project, multiple industrial and academic partners joined forces with two large municipal Swedish hospitals with a goal to analyze, improve and optimize the efficiency and quality of the hospitals' operations and services being provided. One of the identified parameters to be improved was patient waiting time, i.e. the time during which the patient is forced to a passive wait due to a lack of resources of some kind. In the scope of this project, we were aimed to develop an advanced simulation tool which, in addition to beds (see Shahpori et al. [1]), could consider other resources including personnel. For this purpose, predefined standard treatment flowcharts should be correlated with the patient flow data and resource availability data. However, the lack of data with sufficient quality showed to be a big obstacle on the way to achieve the desired goal. First of all, the provided to us predefined standard treatment flowcharts were not general enough to capture a considerable group of patients. The attempt to extract the actual treatment flowcharts for particular diseases using a data+driven approach, similar to the reported in the literature solutions by Zhang et al. [2] and Takagi et al. [3], did not work out due to poor quality and high complexity of the data. Therefore, it was decided to use a visualization tool to better understand the data and, possibly, manually extract the needed information. The conventional analysis tools, such as spreadsheets with standard statistical methods and visualization possibilities and business analytics platforms, employed by the medical staff in the partnering hospitals, turned to be restrictive and impractical for getting the necessary insights from the data. These general-purpose tools are good for providing statistical summaries but they are not capable to present the data from various angles, provide possibilities for interactive discovery of patterns, and give access to details on demand. Therefore, it was decided to develop a data analytics tool that would allow us to conduct an extensive visual explorative data analysis to better understand the data and find potential ways to identify/define the standard treatment flowchart.

In this paper, we present the developed web portal with custom-tailored data visualizations for interactive exploration of patient flows, identifying most visited departments and common transitions, spotting bottlenecks, and excessive waiting times including some hints and estimation of the root cause of the same. In particular, our web portal visualizes several real-world medical concepts, namely a) patient flow, b) hospital logical structure, c) waiting times, and d) episodes of care. In our work, we have managed to create compact yet informative data visualizations that enable an exploratory analysis of the data and provide a bird-eye view over the data while allowing access to details on demand.

II. METHODOLOGY

In the scope of the project, we have arranged recurring visits to the partnering hospitals with the purpose to build background knowledge about the medical domain in general and patient flows in particular. We visited different departments, observing their daily routines, and interviewed staff with various roles and responsibilities. Besides, we arranged interactive workshops with hospital managers, medical staff, and other project stakeholders. During these sessions, we tried to actively engage the participants in creating new and exploring existing ideas of how to make hospitals more efficient.

From the field studies, we have learned that queues of patients usually occur due to lack of shared resources, i.e. medical staff and timely-limited non-human entities (e.g. operation

and treatment rooms, bed departments, medical equipment), and lack of centralized scheduling mechanisms. We concluded that the centerpiece here is the patient-personnel-resources nexus and the target would be to maximize the synergies taking into account all needs, limitations, and quality requirements. The activity to be optimized is the flow of patients, i.e. patients having similar diagnoses and "competing" for the same treatment should follow a well-defined treatment plan. To do so, all the related human resources must also be coordinated accordingly, i.e. personnel needed for each patient visit or stay at the hospital taking into account their working shifts, capacity, and skill set to perform the needed activities. Furthermore, the needed treatment material should be timely arranged and should not pose a limitation if a valid forecast exists.

In line with many authors, see [4], [5], we believe that it is of out-most importance to involve intended users into the development process in order to produce tailored interfaces capable of supporting their domain-specific practices and techniques. Therefore, we relied on the iterative design process. We had 6 iterations during less than 2 years of the project lifetime. At the end of each iteration, we presented the solution to the clinicians, collected their feedback, and refined the design during the next iteration.

III. RELATED LITERATURE

With respect to the discovered literature, several general needs in medical data visualization can be outlined. First of all, authors [4]–[10] point at growing amounts of medical data and express a need for innovative interactive interfaces that would allow systematic yet flexible visual analytics processes by providing data overviews with possibilities of exploration in depth. Second, authors [4], [5], [8], [11] express a need for visualization of patterns and relationships between events for the purpose of discovering cause-and-effect phenomena in medical data.

Temporal relationships and their visualization are heavily investigated topics in the medical domain. The timeline, the predominant metaphor for depicting temporal visualizations in the medical domain [8], usually uses a horizontal axis to express the passage of time and a vertical axis to present variation in time of the variable(s) of interest; point events and intervals are depicted using geometric objects or icons to visualize the chronological sequence of events. This metaphor was first used by Cousins and Kahn [12] who developed an interactive environment that allowed representing complex temporal data as events on an abstract entity called a timeline and enabled users to perform timeline operations such as removing events that do not satisfy some condition, adding new events, and combining two timelines into a single one. Later well-known timeline-based systems are LifeLines [13], [14], a system focusing on visualizing a single patient record, LifeLines2 [15], a system visualizing multiple patient records by stacking instances of single patient records, and LifeFlow [16], EventFlow [17], and Outflow [18] systems that aggregate and provide abstractions of multiple event sequences.

Visualization of different aspects of patient flow data is recognized as an important aspect of medical data presentation and analysis [19]. To plot various quantitative attributes of patient flows, some authors use conventional charts, e.g. see the solution by Jensen et al. [20], others encourage to use custom-tailored visualizations, e.g. see the visualization approaches by Chazard and Beuscart [6], Takagi et al. [3] and Brodbeck et al. [9].

There is only limited work available on the topic of waiting times visualization. Fotiadou [21] concludes that the most informative visual indicator of waiting times is the X percentile, i.e. the time within which X percent of the patients were admitted to the hospital, whereas median and average waiting time indicators, as well as the total size of the queue, were considered useful as complementary information. The author proposes to use speedometer or traffic light metaphors for visualizing the status of waiting times and graphs for the cases when more detailed information about waiting times is needed. Rajabiyazdi et al. [22] show average waiting times as arcs with more detailed information about waiting times presented in scatter plot matrices. In the dashboard by Nagy et al. [23], waiting times are visualized using a dual-axis graph showing average outpatient waiting times using a line and patient volumes using histogram bars; a speedometer-like gauge shows the percentage of patients who had been waiting more than 1 hour. In the dashboard by Franklin et al. [24], the duration of a patient waiting time is indicated by the position of the patient symbol on a timeline; the color of the patient symbol indicates whether a waiting threshold has been exceeded. The dashboard also shows the number of available beds hinting why the patients are waiting.

IV. DATA

In the scope of the project, we were working with the data received from one of the two partnering hospitals. Totally we have received two documents, i.e. an excel document containing historical data about patient visits and an excel document with the scheduling of resources for approximately the same time period. The patient visits data contains records of in- and outpatient visits to the hospital during a time period of approximately 1 year. Each patient visit is stored as a row with 53 columns of attributes. Totally the file contained 75000 rows. The data is anonymized, i.e. there are no personal names, ages, addresses, etc., also patient IDs are encrypted. In the visualization, we used only a few data attributes that were directly related to the project goals. These are the following:

1) **Patient ID**: a unique identifier of a patient. IDs are persistent across the database.
2) **Diagnosis**: diagnosis data is presented by 5 data attributes. This is due to the fact that diagnosis ICD (the International Classification of Disease) codes are grouped into a logical hierarchy with 5 hierarchy levels which categorize the codes according to disease specificity.
3) **Department**: the visited hospital's department is presented with 3 data attributes. This is due to the fact that

hospital departments are arranged into an organizational hierarchical structure that facilitates the management of the hospital. The hierarchy consists of three levels. The actual names of the departments are at the bottom of the hierarchy, the two upper levels containing clinic names and division names group the departments according to certain organizational principles.

4) **Visit type**: each visit in the data is marked with either "new visit" or "revisit". The former means that this visit starts a new episode of care. The latter means that this visit belongs to the ongoing episode of care. Initially, there were other visit types in the data, but we were recommended to ignore such records for simplicity.

5) **Time-related data**: the admission date and time, and the discharge date and time (the latter is only relevant for inpatients).

6) **Operation**: the flag shows whether the inpatient had an operation during the stay at the hospital.

7) **Roentgen**: several attributes specify whether the patient had one or several roentgen scans during the visit.

The excel sheet with resources data contains a table with a daily overview of available/occupied beds in various hospital departments over a period of approximately one year. The departments are presented by their names which match with the departments in the patient visits data. However, not all the departments are listed in the resources document.

A. Data Limitations

In practice, the received data had many limitations. First of all, visits of a patient are not explicitly grouped into episodes of care related to particular diseases. "Visit type" flag allows only to approximately estimate episodes of care. Its main limitation is that, in the case of a revisit, it is not clear which previous visit it is related to. In simple cases, when the patient is visiting the hospital in connection to only one health problem, the revisit is obviously related to the most recent preceding visit. However, if a patient has been visiting the hospital in connection to several co-existing health problems, it is a challenge to separate several ongoing in parallel episodes of care only relying on the "visit type" flag. Furthermore, in the data, every time the patient drops into the emergency department, the visit is classified as a new visit, i.e. a new episode of care is initiated. In the project, we had no other choice rather than splitting patient visits into separate episodes of care relying on "visit type" flag, i.e. starting from a "new visit", we treat all consequent revisits as belonging to a common episode of care until the next "new visit". Theoretically, correct episodes of care could be obtained by performing a more thorough data collection process in the hospital. However, uncertainties would still remain due to the ambiguity of some medical data. We believe that, in any case, for applying data-driven methods on such medical data, manual analysis of data is inevitable to ensure whether the data is reliable.

Second, in the data, there is no information about waiting times, i.e. we do not know how long a patient had been waiting before getting the necessary medical intervention. We solve this problem by calculating a very rough approximation of waiting times, namely, we calculate the time gap between a visit in focus and its previous visit (if there was any). Such an approach allows us to get at least some numeric estimations, calculate more advanced identifiers, e.g. mean times and standard deviations, and identify outliers, i.e. patients with extremely short or long waiting times. The approximated data is good enough for the purpose of demonstrating the capabilities of our web portal, however, it cannot be used for getting a precise picture of waiting times in the hospital. We believe that the only way to get reliable numbers is to properly record it on each patient level.

Finally, some data is occasionally missing, e.g. dates, departments being visited, diagnoses, etc. This sparsity in the data indicates that either the hospital's IT system is not state-of-art or the hospital needs to apply a more thorough approach for recording, storing, and retrieving the data.

V. THE WEB PORTAL

In the scope of the project, we have designed and developed a high-fidelity prototype of a web-based portal that provides highly interactive access to the patient flow data and supports its exploration and interpretation by visually presenting different aspects of the data. Our web portal supports the popular browsing and searching strategy of the Visual Information Seeking Mantra [25], which suggests to provide a high-level overview first, support zoom and filtering, and give access to necessary details on demand. The web portal consists of two interfaces: a) the patient flow view, and b) the episodes of care view. The first interface provides a bird's-eye view of the patient flow in the entire hospital with a possibility to look more closely at each of the departments by expanding their nested departments or by zooming into a particular department for having a closer look at the waiting times or other indicators. The second view provides an opportunity to scrutinize into particular episodes of care, to analyze common transitions and most visited departments. The user can filter the information on both views by various aspects of the data, e.g. by diagnosis or by a time frame. The user can use the two views independently of each other or can use the patient flow view to select the episodes of care which had particular waiting times at a particular department and then inspect them in the episodes of care view.

From the technical perspective, the system consists of a back-end server developed using ASP.NET technology and a front-end developed using HTML and JavaScript; the data is stored as static files on the server side. The interface can be accessed through a modern web browser. In the interface, we employ a combination of customized versions of state-of-the-art information visualization layouts such as treemaps, Sankey diagrams, force layouts, etc. We used an open-source framework D3.js [1] to simplify working with data layouts and animation.
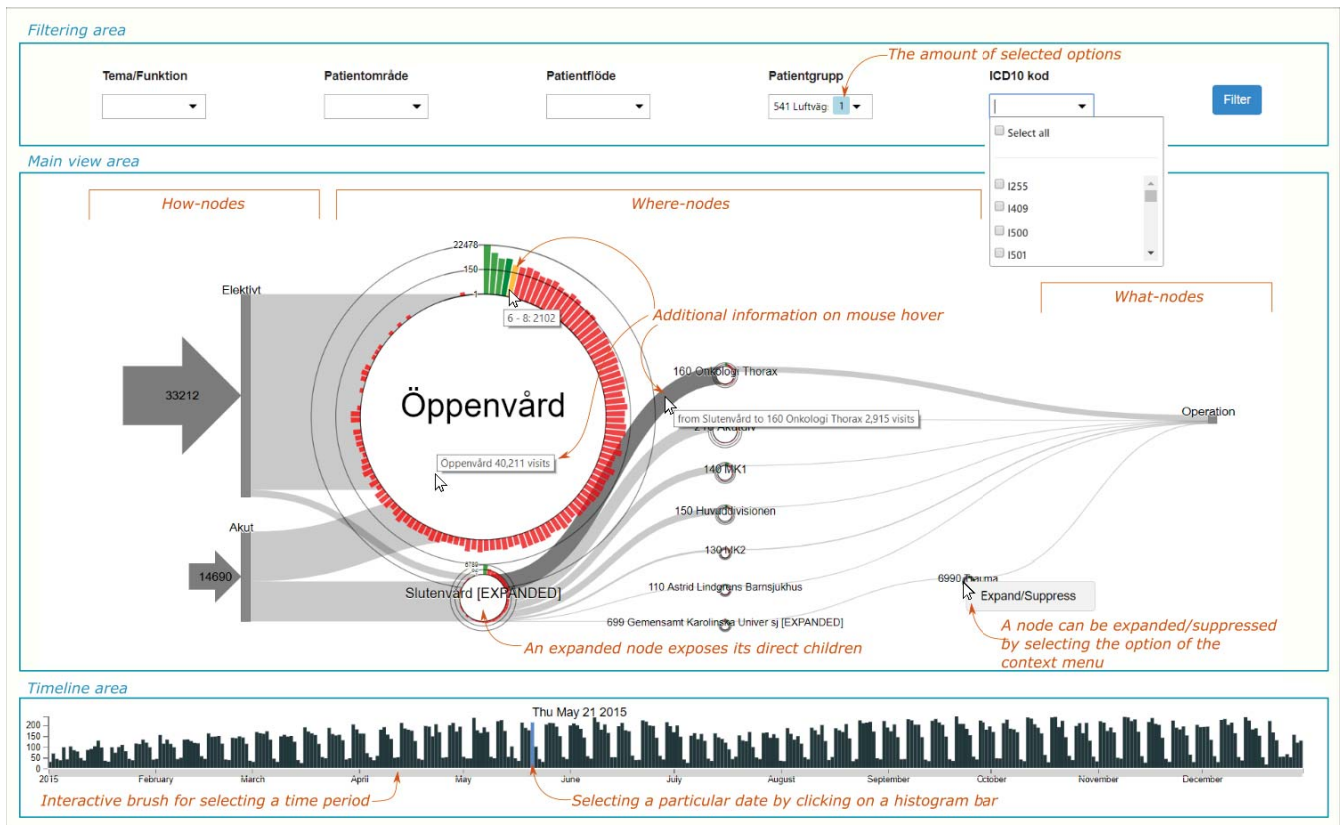
[1] https://d3js.org/

Fig. 1. The patient flow overview.

A. Patient Flow View

The interface, see Figure 1, consists of several areas. The area on the left provides filtering functionality. Filtering can be done by the diagnosis. Filters are implemented as drop-down menus, a drop-down menu per hierarchy level of the ICD diagnosis code hierarchy, totally 5. The drop-down menus allow search for options by typing, selecting multiple values and selecting/deselecting all values at once. The total number of selected options is shown in the header of the drop-down menu. In order to initiate filtering, the user should click on *Filter* button.

At the bottom of the interface, there is a histogram showing the number of patient visits every day during a period of one year. The x-axis depicts the time flow, the y-axis shows the number of people visited on a particular day. One can clearly see various patterns in the periodicity of the visits: daily, weekly, seasonally, etc. The timeline is interactive, the user can select a period using a brush or a date by clicking with the mouse which will result in the patient flow view getting updated to show only the visits that occurred during the selected period of interest.

In the center of the interface is the patient flow view. The view depicts the logical structure of the hospital and the flow of patients visited each of the departments over the selected

period of time. The visualization employs an extended version of Sankey diagram [26]. The extensions are a) custom views of the nodes depending on their types, and b) the ability of the nodes to expand/suppress exposing the hierarchical structure of the hospital.

The first extension was introduced due to the fact that, for the design of this interface, we have applied narrative visualization design principles. This approach is in a way similar to the solution by Zhang et. al [7] who in their design process have oriented towards the definition of a narrative story and visualized the medical data according to the Five W's (who, when, what, where, and why) of journalistic reporting. In our case, we present each patient visit from three aspects:

1) **How** did the patient enter the hospital, i.e. through elective care or through the emergency department.
2) **Where** did the patient go, i.e. which particular department.
3) **What** extra service did the patient receive during the visit, e.g. an operation, roentgen, tomography, etc.

Therefore, in our visualization, we use three types of nodes, i.e. how-nodes, where-nodes, what-nodes. The how-nodes represent possible entry points, i.e. emergency and registry. They are depicted as arrows symbolizing the fact that patients are *coming in* to the hospital. The size of the arrow, as well as

1075

the number inside, shows the number of people entered the hospital through this entry point.

The where-nodes, i.e. hospital departments, are presented as circles. The name of the department is depicted in the center of the circle. The size of the circle is proportional to the number of people who have visited the department during the selected period of time. The choice of using the circle for depicting a where-node was motivated by the fact that the circle is the most efficient 2D-shape in terms of the ratio between its perimeter to the inner area. We wanted to have a large inner area inside the nodes for depicting indicators and other relevant information about the department. In the project, we did not have access to such information, therefore, we left the inner space of each node blank. By clicking on a node, the user will get an enlarged view of the node. The visual transition is achieved by means of smooth animation. This feature is especially useful in case of small-sized nodes.

As mentioned earlier, where-nodes represent the hierarchy of logical structure of the hospital. By default, the interface shows only two where-nodes representing the highest level of the hierarchy, i.e. the primary (öppenvård) and secondary (slutenvård) healthcare wards. The nodes can be expanded or compressed through mouse interaction. When a where-node gets expanded, all its direct child nodes are shown on the UI. When a where-node gets expanded or suppressed, the entire patient flow view get horizontally expanded/shrank, the visual transition is achieved by means of smooth animation.

Around each where-node, there is a histogram depicting waiting times of that department, i.e. the amount of time patients of this department wait to be seen/admitted. The histogram is calculated based on the waiting times of all patients who visited the department in focus during the selected period of time. In the histogram, bars group waiting times into ranges. On the analogy with a clock, the amount of waiting time increases clockwise, i.e. the bars at 00:00 represent the shortest waiting times, the bars at 23:59 PM represent the longest waiting times. The higher is a bar, the greater is the number of patients whose waiting time falls into the range. We use the k-means algorithm to group waiting times into ranges, we use a logarithmic scale to avoid too tall or too short bars. By hovering a bar with the mouse, the user will get a tooltip with the numerical information about the range of the waiting time and the number of patients who have waited for that amount of time. The bars are color-coded based on whether the waiting time they represent is shorter (green), slightly longer (yellow) or exceeds (red) an estimated waiting time threshold for the department in focus. By clicking on a bar, the user is redirected to the episodes of care view. Here all the episodes of care which involved a visit to the department in focus with a waiting time falling into the range of the clicked bar are listed. By scrutinizing into particular episodes of care, the user can analyze the possible reasons for the occurrence of the waiting time.

The what-nodes are depicted as solid rectangles. In the received data, we only had access to one type of extra service, i.e. surgeries.

An edge depicts the number of people visited its target node. It is important to emphasize, that the edges between where-nodes are not showing *transitions* of patients between different departments as this is not the goal of the patient flow visualization, instead, they highlight the *parent-child relations* between the departments. Edges are visually shown as solid gray lines, their thickness is proportional to the number of people who have visited the to-node. By hovering an edge with the mouse, the user will get a tooltip with the number of patients who have visited the target node.

B. Visualization of Episodes of Care

The goal of this view is to provide detailed information about particular episodes of care and to enable the user discovering common patterns in the patient trails in the hospital. The interface, see Figure 2, consists of several areas. The area at the top contains filters. The filters are identical to the filters used in the patient flow view. The looked-up episodes of care are listed on the UI grouped by the patient ID in an expandable list. When clicking on a patient in the list, the row will expand exposing the list of the patient's episodes of care listed by their IDs. Each episode of care has a numeric attribute called relevance. The attribute consists of two numbers. The second number shows the number of visits to the hospital made in the scope of this episode of care, the first number shows how many of these visits are relevant to the search criteria, e.g. have the diagnosis code that the user was filtering by. We show not relevant to the search episodes of care so that, if needed, the user can better restore the original picture of the patient's diseases. A particular episode of care can be selected by checking the checkbox next to its ID. A selected episode of care gets visualized in the area to the right of the list.

We visualize an episode of care using two alternative visualizations, i.e. the timeline view and the transition view. The two approaches are not competing with each other in informativeness, they are rather two alternative presentations of the same data emphasizing different aspects of the data and therefore complementing each other.

The timeline visualization approach is in many respects inspired by the traditional timeline visualization. Similar to the conventional clinical timeline visualizations, we employ a horizontal x-axis for depicting the flow of time and the vertical space above it for depicting data and events in the linear chronological order. The visualization is static, i.e. no animation is involved for depicting the notion of time. Two types of temporal primitives make up the time axis, namely time points and time intervals. We treat an outpatient's visit to the hospital as a time point with the timestamp of the arrival at the hospital. Even though in practice, such visits have a certain duration, this duration has no practical meaning for the purpose of our research, i.e. it has no direct influence on the formation of queues or on common trails. We treat an inpatient's stay at the hospital as a time interval starting at the patient's arrival time and ending at the discharge time. The duration of a patient's stay at the hospital is of interest because, during the stay, shared resources are being consumed which
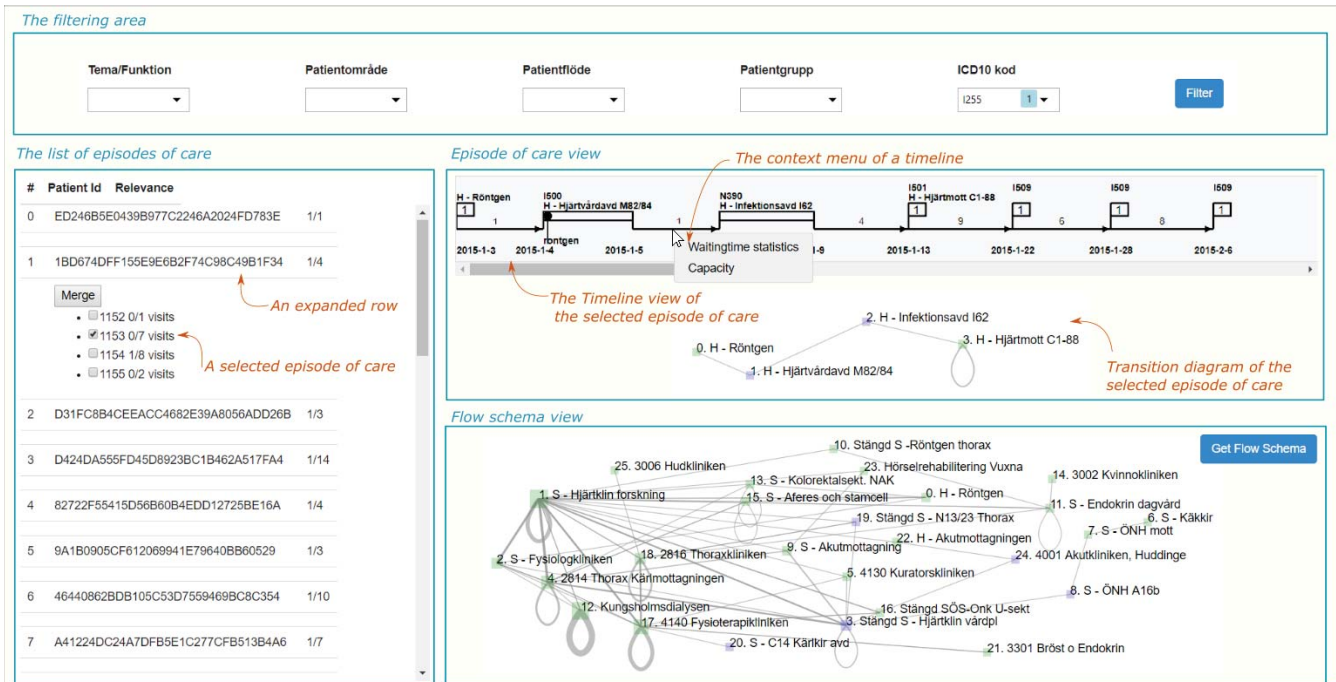
1076

Fig. 2. The episodes of care view.

directly influences the formation of queues. The temporal granularity [27] that we employ in our visualization is fixed to 1 day, i.e. we use a single visual encoding on a timeline to depict all the events that happened on the same day. Such granularity is selected due to fractions of a day do not add much information to the overall picture of waiting times. In addition, suppressing the events that happened on one day into one visual mark on UI reduces visual clutter and the total length of a timeline. The latter is a common timeline problem mentioned by many authors [8], [11], i.e. in case of a long sequence of events, the visualization does not entirely fit on the screen forcing the user to scroll the view horizontally. We distinguish between the behaviors and properties of points and intervals in line with Allen's vision [28]. Due to the data specifics and the selected granularity, we distinguish several types of temporal relations between events, namely *X before Y* and *X during Y*. The latter is related only to intervals, i.e. hospital stays, containing point events, i.e. doctor visits.

We use different visual representations of time points and intervals. A time point is depicted with a rectangle connected to the timeline with a line. The number inside the rectangle shows the number of visits happened on that day. The information about the visit(s), i.e. the name of the visited department and the diagnosis, are shown above the icon. In case of several visits on the same day, the labels are made floating similar to the solution developed by Fekete and Plaisant [29]. This was done in order to overcome the issue of overlapping labels. A floating label automatically places itself in the available space around the icon trying to avoid overlapping with other labels.

An interval is depicted as an oblong rectangle with two lines connecting it to the timeline, one at the beginning of the interval marking the start time, the second at the end marking the end time. The length of the rectangle is proportional to the length of the time interval. The information about the stay, i.e. the name of the hosting department and the diagnosis, are shown above the icon. Nested point events, e.g. surgeries and roentgens, are depicted inside the rectangle as circles. All the labels are made floating to avoid overlapping.

The timeline consists of segments depicted as arrows pointing to the right to emphasize the time period that has passed between every two consequent events. The minimal length of a time period equals the selected temporal granularity which is, as mentioned earlier, 1 day. In the code, we predefined the minimum and maximum possible lengths of the arrow on the UI, the actual length of a particular arrow is proportional to its time period. The length of the time period is depicted in days above the arrow. The start date and end date of the time period, i.e. the dates of two consequent events, are depicted below the arrow, at its start and end accordingly. By right-clicking on an arrow, the user gets a menu of possible actions. One action is to request time statistics, i.e. how long time the transition between the same departments took for other patients. Having selected this action, the user gets a popup window with a box plot depicting minimum, maximum, mean time and standard deviation as well as the value of the current time period (for comparison). The second action is only available for arrows depicting time period preceding a hospitalization. Having selected this action, the user gets a

1077

popup window showing a calendar for every day of the waiting period with numbers of available as well as occupied beds at the department in focus. This information can give the analyst insights into why the waiting occurred.

In contrast to the timeline view, the transition view neglects the time aspect and emphasizes the patient's transitions between various departments of the hospital. Such visualization facilitates spotting patterns in the patient's trails. The visualization consists of nodes, i.e. visited wards, presented as rectangles and connection lines representing transitions between the wards. A connection line in the form of a loop means a consecutive visit to the same department. The departments that have been visited multiple times and the transitions that occurred more than once are visually highlighted proportionally to the number of such occasions, i.e. the rectangles are getting proportionally enlarged and the connection lines get proportionally thicker. By hovering with the mouse over the visual elements, the user gets precise numerical data on demand. In comparison to the timeline view, the visualization remains rather compact regardless of the number of transitions or visited wards. In addition, the nodes of the visualization are movable, therefore, in case of any overlaps or clutter, the user can spread the nodes to make the visualization more readable or rearrange them according to some logical order to highlight the discovered transition patterns.

The interface enables the user to merge several episodes of care of a patient into one. This is handy in case the user believes several episodes of care belong to the treatment of the same disease. Merging several episodes into one will result in getting the treatment history depicted on one timeline, but most importantly will result in a merged transition diagram allowing to see even more trends in the patient's transitions.

Finally, the interface gives the possibility to merge episodes of care of different patients to a common transition diagram. This feature is intended for identifying common trails of patients with a certain disease. To get a reliable picture, a sequence of steps should be taken. First of all, the user should filter all the episodes of care connected to a certain disease/diagnosis. Second, the user should go through all the found episodes of care and pre-select those that they consider relevant or representative to the disease. This is an optional step; however, if one merges all the found cases without pre-selection, the resulting picture will be full of noise. Finally, the user should click *Get Flow Schema* button which will result in the common transition diagram based on the data of all the pre-selected episodes of care. Since the nodes are movable, the user can rearrange the visualization by visually grouping more and less relevant nodes to common clusters.

VI. Preliminary Evaluation

In the scope of the project, we did not have a dedicated evaluation session. Instead, at the end of each iteration, we were demonstrating the tool to the stakeholders and collecting their inputs and requests. Furthermore, at the end of the project, at the urging of the stakeholders, we have uploaded the final version of the tool to a server of the hospital so

that they could play with the tool hands-on without our guidance or supervision. Mostly, the comments were related to requesting additional functionality, e.g. advanced filtering features. The users required minor changes on UI mostly related to renaming textual labels used in the interface. Beside that they did not express any wishes about changing the user interface. We considered it as a positive sign as the users obviously could understand our visual encodings and were generally satisfied with how the data was visualized. The fact that our visualization solution had provided new insights about the hospital workflows was acknowledged several times as a direct input to us or embedded in the discussions. Preliminary evaluation sessions demonstrated that the medical staff of the partnering hospital was excited about trying out the web portal. The clinicians confirmed that they do not have any analogs to the tool as it allows to get insights that can not be gained by studying raw data, or when using conventional visualization and statistical tools. The stakeholders expressed an opinion that the tool has the potential to become a useful and perhaps essential instrument for a modern hospital applying data-driven management approaches.

VII. Discussion and Future Work

The developed web portal is an example of an interactive data visualization tool that enables an exploratory analysis of patient flow data. Our custom-tailored visualizations provide access to various aspects of the hospital's ecosystem, from a high-level overview of patient flows to individual treatments within a single episode of care with just a few clicks. The principles of the Visual Information Seeking Mantra [25] used in our visualizations make them compact yet informative. According to the taxonomy of Bui and Hsu [8], our visualizations are at the highest level of graphical abstraction as they aim to be graphical surrogates of the real-world concepts seen in the clinical environment, i.e. episodes of care, the logical structure of the hospital, and patient flows.

Essentially, the visualization concepts used in our work are not novel, however, their look-and-feel and interactivity make them more modern and elaborated compared to earlier proposed solutions. For example, some interactive features of our timeline, to our knowledge, are unique, i.e. a) visualizing the availability of shared resources during a waiting period, b) allowing operations with the events on a timeline, as well as merging several timelines into one, for enabling the user to outline 'credible data' [30], i.e. the episodes of care that are suitably representative of a disease, to enable productive analysis in the future, and c) showing the timeline visualization together with a transition diagram. Our Sankey diagram stands out compared to the prior solutions due to several reasons. First of all, it visualizes multiple data types, namely numerical, i.e. the number of visits, categorical, i.e. the how-where-what categorization, and hierarchical, i.e. the logical structure of the hospital, and shows relations between them. Second, we adopt hierarchical aggregation to provide levels of detail in a user's analysis. In particular, the advantage of our solution over the prior solutions is in the ability of the nodes to expand

and collapse exposing or hiding the logical structure of the hospital. This approach allows us to provide the user with a compact bird-eye view over the data along with the possibility to drill down into details.

A natural next step will be to validate and refine our prototype by testing it in realistic environments with clinicians. For this purpose, a long-term evaluation test in real settings similar to the setup arranged by Perer et al. in [31] could be desirable. In the absence of the "standard treatment flowchart" for different diagnoses, the next logical step for our work is to proceed with mining of the patterns in order to derive the desired flowchart by utilizing suitable machine learning methods. For patterns mining, a natural way would be to employ Frequent Sequence Mining technique which is a popular tool for finding sets of frequently occurring subsequences from a larger set of temporal event sequences. Reasonable extensions of the algorithm that might be also relevant to our work have been proposed by Perer et al. in [31], [32].

REFERENCES

[1] R. Shahpori, N. Gibney, N. Guebert, C. Hatcher, and D. Zygun, "An on-line dashboard to facilitate monitoring of provincial icu bed occupancy in alberta, canada," *Journal of Hospital Administration*, vol. 3, no. 1, 2014.

[2] Y. Zhang, R. Padman, and N. Patel, "Paving the cowpath: Learning and visualizing clinical pathways from electronic health record data," *Journal of biomedical informatics*, vol. 58, pp. 186–197, 2015.

[3] H. Takagi, K. Misue, and Y. Kanai, "Queuing network model and visualization for the patient flow in the obstetric unit of the university of tsukuba hospital," in *2014 Annual SRII Global Conference*, pp. 147–156, IEEE, 2014.

[4] D. Gotz and D. Borland, "Data-driven healthcare: Challenges and opportunities for interactive visualization," *IEEE computer graphics and applications*, vol. 36, no. 3, pp. 90–96, 2016.

[5] B. Schneiderman, C. Plaisant, and B. Hesse, "Improving health and healthcare with interactive visualization methods," *HCIL Technical Report*, vol. 1, pp. 1–13, 2013.

[6] E. Chazard and R. Beuscart, "Graphical representation of the comprehensive patient flow through the hospital," in *AMIA Annual Symposium Proceedings*, vol. 2007, p. 110, American Medical Informatics Association, 2007.

[7] Z. Zhang, B. Wang, F. Ahmed, I. Ramakrishnan, R. Zhao, A. Viccellio, and K. Mueller, "The five ws for information visualization with application to healthcare informatics," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 11, pp. 1895–1910, 2013.

[8] A. A. Bui and W. Hsu, "Medical data visualization: Toward integrated clinical workstations," in *Medical Imaging Informatics*, pp. 139–193, Springer, 2010.

[9] D. Brodbeck, M. Degen, and A. Walter, "Supporting strategic planning with interactive visualization-a case study of patient flow through a large hospital.," in *HEALTHINF*, pp. 85–93, 2013.

[10] M. Workman, M. F. Lesser, and J. Kim, "An exploratory study of cognitive load in diagnosing patient conditions," *International Journal for Quality in Health Care*, vol. 19, no. 3, pp. 127–133, 2007.

[11] W. Aigner, S. Miksch, W. Müller, H. Schumann, and C. Tominski, "Visualizing time-oriented data—a systematic view," *Computers & Graphics*, vol. 31, no. 3, pp. 401–409, 2007.

[12] S. B. Cousins and M. G. Kahn, "The visual display of temporal information," *Artificial intelligence in medicine*, vol. 3, no. 6, pp. 341–357, 1991.

[13] B. Milash, C. Plaisant, and A. Rose, "Lifelines: visualizing personal histories," in *Conference Companion on Human Factors in Computing Systems*, pp. 392–393, ACM, 1996.

[14] C. Plaisant, R. Mushlin, A. Snyder, J. Li, D. Heller, and B. Shneiderman, "Lifelines: using visualization to enhance navigation and analysis of patient records," in *The Craft of Information Visualization*, pp. 308–312, Elsevier, 2003.

[15] T. D. Wang, C. Plaisant, A. J. Quinn, R. Stanchak, S. Murphy, and B. Shneiderman, "Aligning temporal data by sentinel events: discovering patterns in electronic health records," in *Proceedings of the SIGCHI conference on Human factors in computing systems*, pp. 457–466, ACM, 2008.

[16] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, M. Taieb-Maimon, and B. Shneiderman, "Lifeflow: visualizing an overview of event sequences," in *Proceedings of the SIGCHI conference on human factors in computing systems*, pp. 1747–1756, ACM, 2011.

[17] M. Monroe, R. Lan, H. Lee, C. Plaisant, and B. Shneiderman, "Temporal event sequence simplification," *IEEE transactions on visualization and computer graphics*, vol. 19, no. 12, pp. 2227–2236, 2013.

[18] K. Wongsuphasawat and D. Gotz, "Outflow: Visualizing patient flow by symptoms and outcome," in *IEEE VisWeek Workshop on Visual Analytics in Healthcare, Providence, Rhode Island, USA*, pp. 25–28, American Medical Informatics Association, 2011.

[19] D. Alimov, M. Alexander, N. Gerasim, and H. Mark, "Patient flow visualization and analysis in a complex healthcare institution. (in russian.)," in *Program systems: theory and applications: conference proceedings. Moscow. ISBN 5-94052-066-0*, 2004.

[20] A. Jensen, J. Boyle, and S. Khanna, "Visualising patient flow.," in *HIC*, pp. 77–82, 2012.

[21] A. Fotiadou, "Choosing and visualizing waiting time indicators in diagnostic medical imaging department for different purposes and audiences," *Karolinska Institutet, Stockholm*, 2013.

[22] F. Rajabiyazdi, C. Perin, and S. Carpendale, "West: Visualizing non-emergency surgery waiting times," 2015.

[23] P. G. Nagy, M. J. Warnock, M. Daly, C. Toland, C. D. Meenan, and R. S. Mezrich, "Informatics in radiology: automated web-based graphical dashboard for radiology operational business intelligence," *Radiographics*, vol. 29, no. 7, pp. 1897–1906, 2009.

[24] A. Franklin, S. Gantela, S. Shifarraw, T. R. Johnson, D. J. Robinson, B. R. King, A. M. Mehta, C. L. Maddow, N. R. Hoot, V. Nguyen, *et al.*, "Dashboard visualizations: Supporting real-time throughput decision-making," *Journal of biomedical informatics*, vol. 71, pp. 211–221, 2017.

[25] B. Shneiderman, "The eyes have it: A task by data type taxonomy for information visualizations," in *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pp. 336–343, IEEE, 1996.

[26] P. Riehmann, M. Hanfler, and B. Froehlich, "Interactive sankey diagrams," in *IEEE Symposium on Information Visualization, 2005. INFO-VIS 2005.*, pp. 233–240, IEEE, 2005.

[27] S. B. Cousins, M. G. Kahn, and M. E. Frisse, "The display and manipulation of temporal information," in *Proceedings of the Annual Symposium on Computer Application in Medical Care*, p. 76, American Medical Informatics Association, 1989.

[28] J. F. Allen, "Maintaining knowledge about temporal intervals," in *Readings in qualitative reasoning about physical systems*, pp. 361–372, Elsevier, 1990.

[29] J.-D. Fekete and C. Plaisant, "Excentric labeling: dynamic neighborhood labeling for data visualization," in *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*, pp. 512–519, ACM, 1999.

[30] S. Kandel, J. Heer, C. Plaisant, J. Kennedy, F. van Ham, N. H. Riche, C. Weaver, B. Lee, D. Brodbeck, and P. Buono, "Research directions in data wrangling: Visualizations and transformations for usable and credible data," *Information Visualization*, vol. 10, no. 4, pp. 271–288, 2011.

[31] A. Perer, F. Wang, and J. Hu, "Mining and exploring care pathways from electronic medical records with visual analytics," *Journal of biomedical informatics*, vol. 56, pp. 369–378, 2015.

[32] A. Perer and F. Wang, "Frequence: Interactive mining and visualization of temporal frequent event sequences," in *Proceedings of the 19th international conference on Intelligent User Interfaces*, pp. 153–162, ACM, 2014.

[2]https://www.vinnova.se/en/

1079