Using Dashboard Networks to Visualize Multiple Patient Histories: A Design Study on Post-operative Prostate Cancer

Jürgen Bernard, *Member, IEEE,* David Sessler, Jörn Kohlhammer, *Member, IEEE,* and Roy A. Ruddle

Abstract—In this design study, we present a visualization technique that segments patients' histories instead of treating them as raw event sequences, aggregates the segments using criteria such as the whole history or treatment combinations, and then visualizes the aggregated segments as static dashboards that are arranged in a dashboard network to show longitudinal changes. The static dashboards were developed in nine iterations, to show 15 important attributes from the patients' histories. The final design was evaluated with five non-experts, five visualization experts and four medical experts, who successfully used it to gain an overview of a 2,000 patient dataset, and to make observations about longitudinal changes and differences between two cohorts. The research represents a step-change in the detail of large-scale data that may be successfully visualized using dashboards, and provides guidance about how the approach may be generalized.

Index Terms—Information Visualization, Visual Analytics, Multivariate Data Visualization, Electronic Health Care Records, Medical Data Analysis, Prostate Cancer Disease, Design Study, User Study, Evaluation, Static Dashboard, Dashboard Network.

1 Introduction

Visualization has the potential to become an integral part of medical health care. For that, the visualization of patient histories from electronic health record data is one of the central subjects of interest, and can support medical research, clinical treatment, and communication between medical experts (physicians) and patients [1]. However, real-world patient histories often provide a wealth of information about multiple attributes about patients, e.g., outcome variables and treatment conditions [2]. It follows that a large amount of data is sometimes needed to reason about the well-being and treatment decisions of patients, making the visualization of detailed patient histories a challenging task.

Two current approaches are as follows. The *visualization of single patient histories* has been applied in various applications [3], supporting tasks for different stakeholders [4]. Visualization techniques can show multiple attributes of specific patients in a single diagram. However, the emphasis on details means that these visualizations do not scale well, which makes it tedious to compare multiple patients. To address this, a second visualization approach aims to let users see the *event sequences of multiple patients* at a glance. The strategy of this approach is to abstract complex histories into different event categories or discrete states [5]. Thus, every patient history is represented as a sequence of symbols, which can be visually aggregated to show multiple patients. However, given the fact that every patient has a unique history, this approach involves the loss of a lot of potentially valuable information.

Showing both *multiple patients* and *multiple patient attributes* at the same time remains a research challenge. If suitable visualization techniques could be developed then users would be able to gain an overview of large sets of patient histories and discover details about specific patient attributes. Interesting relations in the data could be identified, e.g., that many patients sharing specific properties in a pathology report will develop similarly in their follow-up. In addition, such visualizations would allow the comparison of different patient cohorts, which would help medical experts to assess specific causes and outcomes.

The present paper describes the design and evaluation of a technique for visualizing patient histories, using post-operative prostate cancer as an example. Overall this paper makes three primary contributions. First, we show that by segmenting patients' histories, instead of treating them as raw event sequences, segments can be aggregated using a variety of criteria (e.g., whole history vs. treatment combinations) and visualized as a compact *static dashboard* that depicts a broad range of treatments and outcome variables. Second, we demonstrate that the static dashboards may be arranged as *dashboard networks* that allow users to make a wide variety of observations about longitudinal and cross-cohort changes in patients' histories. Third, we show that our visualization technique is effective for users who range from medical experts and visualization experts to people who are not expert in either field.

The following sections describe the background of the research, the iterative design process, and then a user evaluation. The evaluation was conducted with three groups of users: non-experts, visualization experts, and medical experts.

2 Related Work

We structure the related work by the class of visualization techniques that was applied, which in many cases depends on the analytical task to be supported. Overall, we follow the distinction proposed by Rind et al. [3] where 14 visualization systems for electronic health records were compared for clinical research or

- *J. Bernard, D.Sessler, and J. Kohlhammer are with TU Darmstadt, Darmstadt, Germany.*
  E-mail: see http://www.gris.tu-darmstadt.de/home/index.en.htm
- *R. Ruddle is with University of Leeds, Leeds, UK.*
- *J. Bernard was with, J. Kohlhammer is with Fraunhofer IGD, Darmstadt, Germany.*

Manuscript received Month xx, 20xx; revised Month xx, 20xx.

practice. First, we review approaches where single patient histories were visualized with an emphasis on techniques including multiple patient attributes. Second, we discuss systems that gave an overview of multiple sequences, which typically were represented as a set of event types or visual symbols. Extending the survey of Rind et al., we summarize visualization techniques for multivariate data with an emphasis on medical data visualization.

2.1 Visualization of Single Patient Histories

Some early techniques for displaying graphical summaries of status of patients were presented by Powsner and Tufte [6], [7] over 20 years ago. That research was complemented by LifeLines [8], which was an interactive tool for visualizing patient histories and used a variety of visual encodings to show longitudinal health record data.

Other tools have been developed more recently. One used the color semantics of a traffic light to indicate patient well-being over time [9]. Bernard et al. [10] also used a traffic light color scheme to combine the progression of the PSA hormone value with multiple clinical and histological dimensions in a tool that allowed users to interactively define cohorts. In a follow-up approach, the authors optimized the visualization of patient histories for an active learning approach for the assessment of patient well-being [11]. Although, these tools allow the visualization of multivariate information about single patients over time, the visualization of multiple patient histories in a compact way, be it juxtaposed or superimposed [12], remains the subject of current research.

2.2 Visualization of Event Sequences of Patients

To visualize the histories of multiple patients, the data is often simplified by either abstracting events to a set of states or aggregating identical events and sequences into higher-level sequences. All of these approaches increase visual scalability, and example tools are LifeLines2 [13], Outflow [14], CareFlow [15], DecisionFlow [16], LifeFlow [17], and EventFlow [18], [19].

A core analysis task in LifeLines2 [13] is the visualization and temporal comparison of multiple records that have been abstracted to event sequences. We take up this idea and apply it to multivariate states of multiple patents using a suite of charts for different patient attributes. Similarly, the LifeFlow approach [17] converts histories to a tree-based data structure that is used for the visualization of simplified sequences. While we also apply aggregation, our visualizations combine different graphics techniques for the visual aggregation of patient states.

EventFlow [19] allows users to adopt filtering and substitution strategies to simplify data. StratomeX enables the interactive editing of patient groups and provides visual feedback for relations across different data sources [20]. Similarly, TimeSpan [21] visualizes temporal event-sequence data and multi-dimensional data for individual patients. It provides matrix representations and reordering techniques for exploring patterns and new hypotheses in data. COQUITO [22] uses networks with abstracted temporal information to show patient cohorts as the result of interactive temporal queries. CoCo [23] applies visual analytics and statistical testing for the visual comparison of cohorts, based on an abstraction of patient histories to event sequences. EventAction [24] also builds upon statistical testing, while emphasizing the combination of summaries of attributes and event sequences. As such, CoCo and EventAction are examples which can be combined with the static dashboard networks presented in this work to facilitate the visual comparison of differentiating attribute values.

To extend the functionality of approaches based on event sequence visualization, we want to visualize multiple patient attributes. The combination of being able to visualize both multiple event sequences and multiple attributes of patients is central to the requirements of medical experts and the motivating challenge for the visualization design of the present research.

2.3 Visualization of Multivariate Attribute Sets

Our visualization design relies on design principles for the visualization of multivariate data. According to the medical experts (prostate cancer clinicians) with whom we collaborated, 15 attributes provide information that is important in prostate cancer health care. Those attributes are of a variety of data types (numerical, ordinal, categorical, and binary), which is why visualization techniques that were designed for numerical attributes (e.g., parallel coordinates [25], RadViz [26], pixel displays [27], or scatterplot matrices [28]) are not appropriate. Therefore, we base our static dashboard on a class of techniques for the visualization of multivariate data with different attribute types (mixed data), using assemblies of small visual objects (charts) to represent individual attributes. For related reviews, we refer readers to the works of Fuchs [29], Borgo et al. [30], and Loorak et al. [31].

The design of intuitive and useful static dashboards for multiple attributes is a non-trivial endeavor, because designers have to consider the characteristics of the data, the intended analytical tasks, and the requirements and preferences of users (e.g., semantics). In addition, designers must be aware of perceptual principles of visual variables, gestalt psychology, and semiology [32], [33], [34], [35]. If designed appropriately then such visualizations show complex data and relationships in a compact and intuitive way, and some examples from the bio-medical domain are as follows.

Similar to our approach, Maguire et al. present a visualization describing progress over time [36] to visualize workflows of biological experiments. Horn et al. present a so-called structured metaphor graphic object for the identification of multiple parameters in the context of intensive care units [37]. We build upon that idea to assemble multiple attributes in a compact and structured way. Other techniques visualize human body configurations, allowing the analysis motion over time, e.g., to assess health recovery. Along these lines, the FuryExplorer approach [38] uses a stick figure metaphor for the assessment of health recovery. Similar to our static dashboard, the visualizations can represent multiple objects, conveying a notion of the variance. An alternative idea is the use of natural objects to encode abstract data. With DICON, Cao et al. [39] present visualizations representing distributions of multivariate data, emphasizing cluster exploration as an important analysis task.

In summary, some of the above approaches focus on the visualization of single patients, while others visualize multiple event sequences in a more abstracted way. These two general choices either allow the visualization of dozens of attributes for few patient histories, or the visualization of a few attributes for hundreds or thousands of patient histories. However, the challenge of visualizing multiple attributes for many patient histories remains unsolved. For this, we consider that an assembly of carefully designed visual objects is well-suited.

3 Application Background

The present research combines a design study methodology [40], [41] with the application area of post-operative prostate cancer to develop methods for visualizing multi-attribute patient histories. We have been collaborating with four medical experts at the Department of Prostate Cancer at the Universitätsklinikum Hamburg-Eppendorf (UKE), who take care of several thousand patients per year and perform over 2,000 surgical procedures per year. This collaboration involved bi-monthly meetings with two of the medical experts, and two additional medical experts who gave feedback during the design study. The following sections introduce prostate cancer treatment, describe requirements for visualizations of multi-attribute patient histories, and characterize the real-world dataset that we used to design and evaluate the visualizations.

3.1 Treatment of Post-Operative Prostate Cancer

Prostate cancer is a disease that affects millions of men all over the world. Clinics gather diverse information about patients in the form of electronic health records, including demographic data, blood and histological samples, clinical data and surveys of follow-up care. One of the medical experts' primary goals is to improve the accuracy of early prognoses, e.g., by analyzing diagnostic information. For this, much effort is spent on the stratification of patients into cohorts to allow cohort histories to be compared, so that medical experts can improve treatment quality and avoid treatments that are harmful or unnecessary.

One of the diagnostic indicators (see Table 1) is the progression of the PSA hormone (a prostate-specific antigene enzyme). Another important biological indicator is the occurrence of metastases, which implies a bad prognosis for a patient and typically occurs in late stages of the disease. Additional information about the type of carcinoma can be gained from histopathology variables that are assessed by the pathologist post-operatively and only once: the *Gleason* score, information about affected lymph nodes (*pN-status*) and the tumor class (*pT-status*). Four of the most common treatments are surgery (*OP*), radiation therapy (*RTX)*, hormone therapy (*HT*), and chemotherapy (*CHT*). These treatments have local (*OP*, *RTX*) or global (*HT*, *CHT*) implications on the human body. Additional outcome variables are gathered to quantify treatment success. The relapse of patients after surgery (biochemical recurrence, *BCR*) is one such an indicator, which is often assessed with the PSA progression, and another is information about death (death of disease; *DOD*).

3.2 Requirements

UKE currently uses two visualization methods. The first is paper print outs of single "patient plots" with multiple attributes, to identify and discuss differences between specific patients. The second is a visual analytics tool that allows medical experts to stratify patient cohorts and correlate cohorts with static attributes [10]. However, several other functionalities remain unsupported. These can be structured into the following three groups of requirements:

1) Giving overviews: Providing an overview of a group of patients for strategic planning with hospital managers or for communicating potential treatment outcomes during patient consultation.
2) Presenting longitudinal changes: Patterns of disease progression and treatment outcomes for use in clinical team meetings or for showing temporal patterns in patient histories.
3) Comparing cohorts: Complementing statistical findings with visualizations that summarize cross-cohort patterns of patient histories in the medical experts' scientific papers.

The complexity of diseases such as prostate cancer means that, to be useful, visualizations need to be capable of presenting multiple attributes of patient histories. For some use cases it is sufficient to present a segment of a set of histories (e.g., for giving an overview), others benefit from showing changes as a disease progresses, and the remainder will typically involve both disease progression and an understanding of differences between patient cohorts (as in the third group above). The audience for the visualizations ranges from people with little knowledge of a specific disease over medical experts to clinical teams that are comprised of different user groups (e.g., clinicians and nursing staff).

3.3 Data Characterization and Abstraction

This section describes the data and its abstraction. We provide details about the data collection as well as the characteristics of patient histories and patient attributes in Section 3.3.1. Next, we present information about data processing, i.e., how we went from patient histories to the data used in dashboard networks. The two core steps in this process are the segmentation of patient histories (Section 3.3.2) and the downstream aggregation of patient segments (Section 3.3.3).

3.3.1 Patient Histories and Attributes

We had access to a pseudonymized portion of the data that contained the histories of patients who had had surgery (*OP*). In addition, all of the patients in this dataset had a relapse after surgery (*BCR*) and received a second type of treatment. The medical experts call such a collection of patients a 'negative-selection', though a relapse is fortunately not common.

The data collection contained almost 2,000 histories, which is one of the largest collections for prostate cancer patients in Europe. The dataset is important to the medical experts for both research and practice, because the dataset underpins data-driven research and is increasingly used to provide scientific evidence for medical hypotheses. During hypothesis generation and validation, one fundamental analysis task is the identification and communication of specific properties of individual patient cohorts. Our visualization design and user evaluation is targeted towards the visual comparison of histories of different patient cohorts, and thus match that task.

For medical experts, the choice of attributes, their abstraction, and visual representation has a significant influence on the usefulness of a visualization for a given analytical task. Based on our previous discussions with UKE experts, we selected 15 attributes describing the conditions from both the biological and the treatment perspective. In this way, we preserved the details of individual patients which can be highly specific since every patient history is unique. Four of the outcome variables reflect the histology report that was gathered after surgery (*pT-status*, *pN-status*, as well as two histological *Gleason* scores). Patients in the dataset had up to four different treatments, leading to four treatment attributes (surgery, radiation therapy, hormone therapy, and chemotherapy). Moreover we include four outcome variables occurring in the course of the post-surgery observation (*OK* after surgery, relapse, metastases, death of disease). The four remaining attributes describe the progression of the PSA hormone (*pre-OP* PSA, PSA at segment *start*, PSA at segment *end*, PSA

Attribute	Description	Att. Type	Value Domain	Visual Enc.
PSA Pre-OP	The value of the PSA hormone is a most important biological indicator for prostate cancer. Last measurement of the PSA hormone before OP (surgery).	Numerical	$[0;200.0[\frac{ng}{ml}$ $median: 8.7$	
PSA at segment start	Value of the PSA hormone at start of the visualized temporal segment.	Numerical	$[0;4,472.0[\frac{ng}{ml}$ $median: 0.21$	
PSA at segment end	Value of the PSA hormone at the end of the visualized temporal segment. Used to assess the progression of the PSA value within the segment interval.	Numerical	$[0;4,472.0[\frac{ng}{ml}$ $median: 0.26$	
PSA trend in segment	Trend of the PSA hormone value within the duration of a temporal segment. Depicts the delta between the PSA values in a relative way.	Numerical	$]-70.3;4381.68[\frac{ng}{ml}$ $median: 0.0$	
pT-status	Histopathological status of the tumor, obtained by pathology report (pT - pathological tumor). Used for prognoses about the progress of the disease.	Ordinal	$pt2a, pt2b, pt2c, pt3a$ $pt2b, pt3c, pt4$	
pN-status	Histopathological characteristics of the tumor, obtained by pathology report. Used for prognoses about the progress of the disease.	Ordinal	$NX, N0, N1$	
Gleason Score	Histopathological characteristics of the tumor, obtained by pathology report. Used for prognoses about the progress of the disease.	Ordinal	$2+3, 3+2, 3+3, 3+4, 3+5,$ $4+3, 4+4, 4+5, 5+3, 5+4$	
OK after OP	Outcome variable (biological condition) reflecting a good condition after OP. Is active until relapse or metastases.	Boolean	$[false, true]$	
BCR	Outcome variable (biological condition) describing the biochemical recurrence (relapse) of the tumor. Used for prognoses about the progress of the disease.	Boolean	$[false, true]$	
Metastases	Outcome variable (biological condition) describing whether a patient got metastases. Metastases are the most severe biological indicator in the dataset.	Boolean	$[false, true]$	
DOD	Biological end point when a patient died of disease. Often used for cause-effect analyses.	Boolean	$[false, true]$	
OP	Treatment attribute reflecting whether a patient had surgery. In the dataset all patents received OP.	Boolean	$[false, true]$	
RTX	Treatment attribute reflecting whether a patient got radiation therapy. Local treatment, only affecting the prostate area.	Boolean	$[false, true]$	
HT	Treatment attribute reflecting whether a patient had hormone therapy. Systemic treatment, affecting the whole human body.	Boolean	$[false, true]$	
CHT	Treatment attribute reflecting whether a patient got chemotherapy. Often applied in severe situations, e.g., when metastases are detected.	Boolean	$[false, true]$	

TABLE 1
Overview of 15 most relevant attributes describing the condition of a patient. We provide individual visual encodings for each attribute. All encodings comply with the requirement of showing information either of single patients or of groups of patients (clusters). When a group of patients is visualized the dimensionality to be depicted increases from 16 (Gleason is 2D) to 19 (encoding of variation). For boolean attributes, such as OP, the pie represents a ratio patients having a distinct attribute. Gleason Score is special as it has two independent integer dimensions.

trend). Table 1 provides an overview of the 15 attributes, their characteristics, and their visual encodings.

3.3.2 Segmentation of Patient Histories

In this work, we refer to *segmentation* as the division of complex temporal data into meaningful time-ordered parts [42]. Temporal segments of multivariate data allow data visualization and analysis in finer granularity. The visualization of individual segments within patient histories is directly supported by the static dashboard visualization technique. The segmentation routine that was used in the present research is the result of a design study that we conducted with medical experts [11]. At the experts' recommendation, each segment of a patient history keeps most information constant so that each new segment signifies a state change in the history.

The segmentation routine worked by traversing a patient's history with a given step size and a maximum window length (one week and six months, respectively, in our case), using a sliding window approach. A new segment was created if at least one of the following two criteria was met. First, the *PSA value* changed, a new treatment occurred (*OP*, *RTX*, *HT*, or *CHT*), or a new outcome occurred (*OK*, *BCR*, *Metastases*, or *DOD*; see Table 1). As a result, every change in a patient history triggers the routine to create a new segment. Please note that the three remaining attributes *pT-status*, *pN-status*, and *Gleason* always change with *OP* (histology report). With the first criterion the length of a segment depends on the aggressiveness of the patient's tumors, but is typically a few months. However, in situations where a patient has a severe condition, there may be rapid changes in the history so each temporal segment covers a shorter period of time. The second segmentation criterion is activated when the traversed segment reaches the maximum length without attribute changes. This criterion allows slower disease progression to be segmented, by creating multiple states for a constant state. A state change in the numerical value of the *PSA* hormone leads to different values at the *segment start* and *segment end*, which indicates the *PSA trend*.

Overall, the routine created 10,485 segments for the 2,000 patients. Typically, a patient history yielded three to seven segments.

3.3.3 Aggregation of Segments

We use the term *aggregation* to refer to the process by which segments were grouped for display in a given static dashboard. The

medical experts proposed several possible aggregation criteria, including the number of treatments, patient outcomes and the value of the PSA hormone, and other criteria could also be applied (see Section 6.4). In the present paper, two criteria are illustrated and were used in the user evaluation (see Section 6). The first was simply to aggregate all 10,485 segments together, to provide an overview of all of the patient histories in one static dashboard (Figure 1). The second was to divide patients into cohorts, and then aggregate each cohort's segments according to the combination of treatments (*OP*, *RTX*, *HT*, or *CHT*) that had been provided. When the resulting static dashboards were assembled into a dashboard network, the number of treatments increases from left to right, and different treatment combinations cause a network to branch (see Figures 3 and 4 for the *pt4* and *pt2c* cohorts, respectively).

Once segments have been created (see Section 3.3.2), the aggregated data to be displayed in each static dashboard is calculated using the following two steps. In the first step, the aggregation routine creates a number of bins, with the specification of bins depending on the aggregation criterion that was chosen. In the second step, segments are assigned to appropriate bins.

The criterion for Figure 1 is that there is a single bin for all segments, whereas in Figures 3 and 4 the aggregation routine creates a bin for every treatment combination. In general, if the criterion is based on a categorical attribute, we recommend that categorical attributes are binned so that the most frequent categories are assigned to individual bins, and remaining categories are grouped together [43]. For numerical attributes (e.g., *PSA*), the bins are defined according to standard binning variants which can be domain-preserving (the default in the present paper), frequency-preserving, or be based on a goodness-of-fit measure [44], [45].

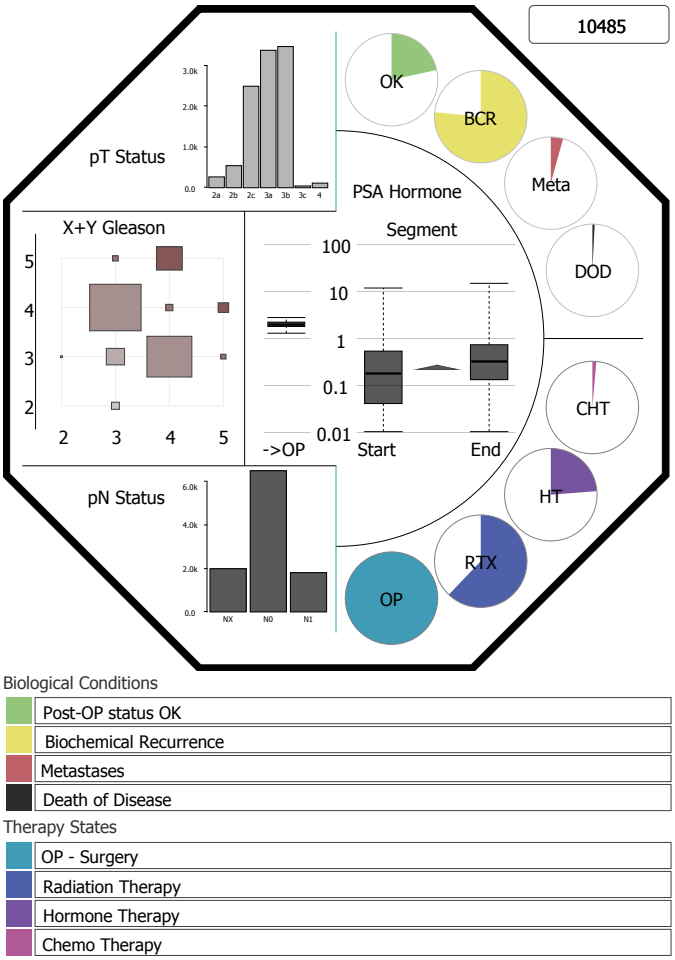

Fig. 1. Patient segment visualization showing detailed information about 10,485 patient segments at a glance. In this example the visualization provides an overview of all patient history segments of the dataset. Depending on the available display space the visualization shows explanatory labels for every encoded attribute. Analysts can assess the distribution of values in detail including label information. In addition, an enlarged visualization (referred to as a view [46]) helps to lookup attribute information for smaller visualizations and thus can serve as a legend in a multiple-views application.

4 Visualization Design

This section describes the design of the dashboard networks, which we developed to visualize multiple attributes in patient histories. The overall challenge was to create a design that allowed users to gain an overview of a set of patient histories (e.g., Figure 1), assess changes that took place over time in a given cohort (e.g., Figure 3), and compare two cohorts (e.g., Figure 3 vs. Figure 4). The following sections describe the design concept, the iterative design process, and four key aspects of the design: the use of color, visual aggregation of segments, chart types and structure.

4.1 Design Concept

The underlying concept was based on three levels of detail. At the lowest level were the charts that presented each attribute in the patient histories. At an intermediate level, those charts were integrated into a dashboard that showed all of the attributes for either the whole history of patients (see Figure 1) or a given segment/cohort of patient histories (see Section 3.3). For this intermediate level we use the term *static dashboard* because users could not interact to change the charts, unlike the dashboards that are commonly used in business intelligence applications. At the highest level, the static dashboards were displayed as small multiples in a *dashboard network* to show each segment in the histories of a patient cohort (see Figures 3 and 4).

Each static dashboard was self-contained, meaning that it summarized a particular state in the patient histories by showing the attribute values. This meant that users could compare any two temporal segments of the patients' histories by looking at the corresponding static dashboards. An alternative approach would have been for each static dashboard to show the changes in the attributes since the previous static dashboard, but that would have made it more difficult for users to compare segments that were either not adjacent in a dashboard network or that are related to different cohorts.

Making the dashboards static meant that users would be able to make observations with the glance of an eye, instead having to interact and rely on their memory of previous views of the data. Static dashboards may also be printed out and are well-suited for use in meetings (e.g., to review the quality of care and patient outcomes in a treatment unit).

4.2 Iterative Design Process

The static dashboard design went through nine main iterations (see Figure 2), during which we interviewed medical experts, tested design alternatives, and held critical discussions with visualization experts. The data attributes that were encoded in the final iteration are described in Table 1, and a summary of the iterations is as follows.

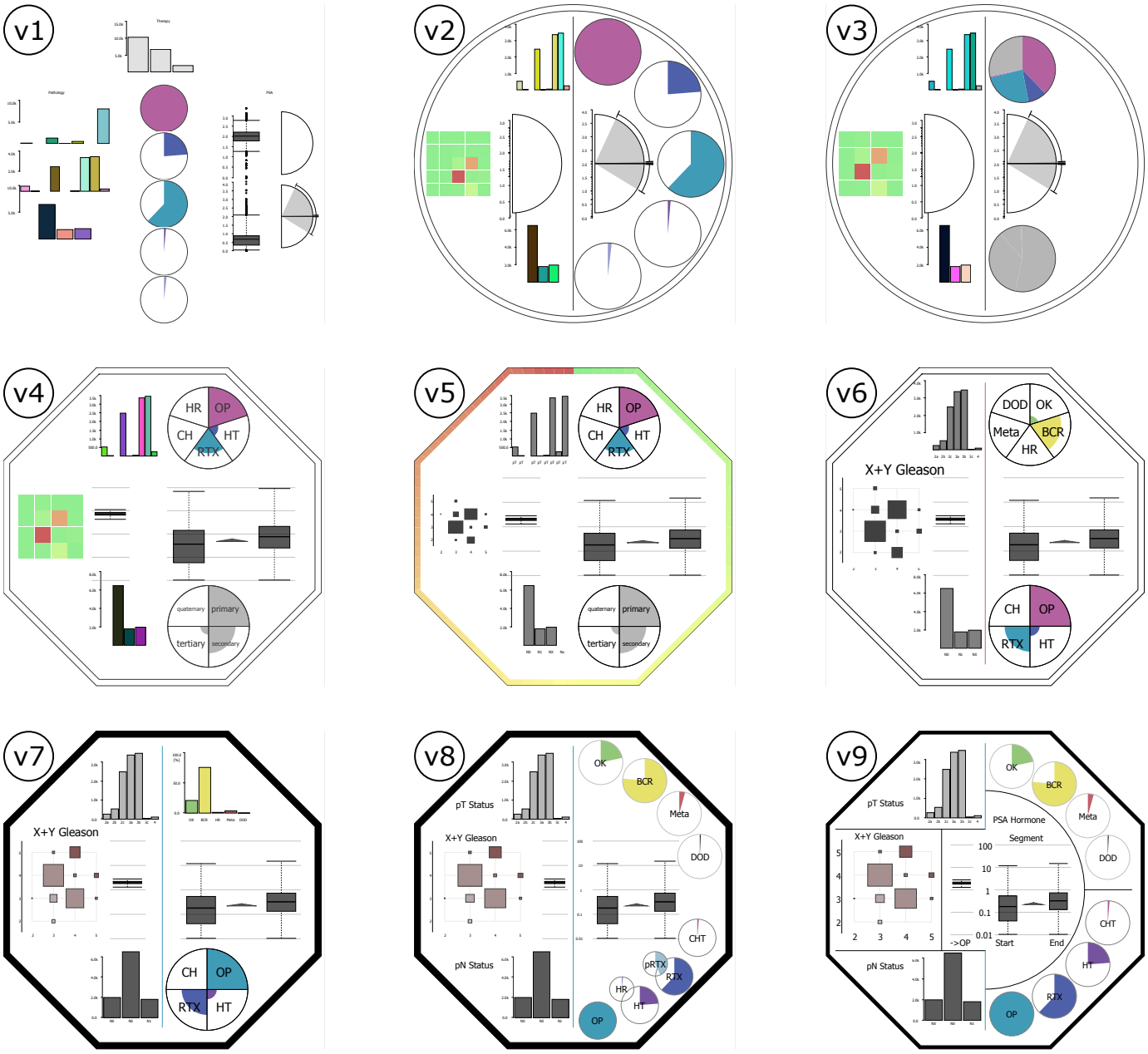

Fig. 2. Evolution of the visualization design. Basically, nine core iterations mark the design process each posing an improved result. Interviews, observations, discourse, and informal tests helped to gain valuable feedback in the course of the process. In summary, crucial factors for improving the visualization were orderedness, attribute grouping, separation, symmetry, and a selective use of color.

$\mathbf{v}_1$ showed four treatment variables (*OP*, *RTX*, *HT* & *CHT*) and an outcome variable (resistance to hormone therapy; *HR*) as pie charts, the *pT-status*, the *pN-status* and the sum of *Gleason* scores as colored bar charts, the number of patients who had undergone from one to four treatments (NB: these are labeled as primary, secondary, etc. in $\mathbf{v}_4$ and $\mathbf{v}_5$) as a grayscale bar chart, and the PSA values and changes as box plots and using a speedometer metaphor, respectively. In retrospect, this first design was rather rudimentary.

$\mathbf{v}_2$ added a circular border to separate each static dashboard within a dashboard network, and included the Gleason scores as a heat map. Other charts were moved to improve the static dashboard's structure (for details, see Section 4.6).

$\mathbf{v}_3$ condensed all of the treatment variables into a single pie chart, and the number of treatments into another pie chart. However, that was not successful because it meant that the pie chart showed the relative frequency of each treatment rather than the proportion of a cohort who had had each treatment (or number of treatments, in the case of the second pie chart). This iteration also separated the pre- and post-surgery attributes by a vertical line for the first time.

$\mathbf{v}_4$ involved three substantial changes. First, the border was changed to an octagon because that allowed the dashboard networks to be more compact and, in our opinion, is more aesthetic than a square border. Second, the pie charts were changed so that the radius of the shaded part of each segment showed the proportion of a cohort who had each had a given number of treatments. Third, the design of the PSA charts was changed after

user tests indicated that the speedometer was not intuitive. The new design used box plots to show the distribution of the PSA values pre-operatively, the PSA values at the start and end of the segment, and the PSA trend during the segment (up or down). The new design met with the approval of the medical experts.

$\mathbf{v}_5$ also involved three changes. Redundant coloring was removed from the bar charts, the Gleason score heat map was changed to use size encoding to further reduce the use of color, but color was added to the octagonal border to show the distribution of patient well-being scores [11]. The latter change was rejected after informal user feedback indicated that the scores were difficult to understand and the users reported "restlessness" in a border element that should be a simple outline.

$\mathbf{v}_6$ replaced data about the number of treatments with outcome variables (*OK*, *BCR*, *HR*, *Meta* and *DOD*), complementing the treatment variables This enabled the experts to conduct cause-and-effect analyses with the dashboard networks for the first time.

$\mathbf{v}_7$ was used to investigate the usage of bar charts to depict outcome variables. However, users tried to identify relations between the three bar charts, instead of relating outcome variables with treatment variables, so that design change was dropped.

$\mathbf{v}_8$ involved three substantial changes. The first was to revert back to showing one attribute in each pie chart, providing simple charts that should be understandable by most users. The second was to position the treatment and outcome variable pie charts so that they were in order of increasing severity. The third was to add pie charts for two treatment-related attributes (*pRTX*: end of radiation, *HR*: hormone refractory), but feedback from the medical experts indicated that the additions were not welcome.

$\mathbf{v}_9$ added lines to divide the static dashboard into regions that each contained the chart(s) for a set of attributes, which was explicitly welcomed by some participants in the user evaluation. The design was also given some final polishing, such as the optimization of label sizes.

4.3 The Use of Color

Color is used to discriminate the four treatment variables, and the four outcome variables. To some degree, this use of color is redundant since the variables can be distinguished by their different positions in a static dashboard. However, interviews with visualization experts suggested that coloring should be included in the approach for three reasons:

- The lookup of colored pies is pre-attentively perceivable
- Color eases visual comparison for small visualizations
- The colors have a semantic meaning

The colors were chosen with input from the medical experts about color semantics of the domain. However, we did equalize the brightness and saturation of the colors, following input from an expert in perception and cognition. The four outcome variables follow the idea of a traffic light metaphor (plus black for *DOD*), according to the severity of the disease. A consequence of the metaphor and the domain-specific colors was that both Meta and CHT used a reddish color, but as both are severe conditions the semantic overlap was considered acceptable. Overall, the colors seemed to be correctly interpreted by participants in the user evaluation. However, the colors are not colorblind safe, which is a possible subject for future work. One option is to display the four biological conditions in different brightnesses of a given hue, with a second hue used for the four therapy states, and a second option is to use a grayscale scheme because the labels (*OP*, *RTX*, etc.) allow each chart to be identified.

In addition to using color for the treatment and outcome variables, we used a gray to dark red colormap to emphasize the severity of Gleason scores, which increases from the lower left to the upper right of the 2D grid. This color encoding supports the visual comparison of Gleason scores.

4.4 Visual Aggregation of Segments

One challenge in the design was showing multivariate information for multiple segments in a single dashboard. The number of segments visualized with a single static dashboard depends on the aggregation strategy applied to assign segments to groups (see Section 3.3.3). For the user evaluation in the present research the number of segments in a group ranged from eight (the right hand static dashboard in Figure 3 to 10,485 (see Figure 1). We addressed the challenge by choosing types of charts that are capable of showing either a portion of items (the pie charts) or a distribution of values (the other charts). Details are provided in the next section.

4.5 Chart Types

A key challenge that we faced was to design a static dashboard that showed many attributes of patients' histories, and a strength of the final design ($\mathbf{v}_9$) was its use of four different chart types (pie charts, bar charts, box plots and a heat map), which helped the groups of attributes to stand out from each other. This section reflects on those choices of chart.

Box plots were used for the PSA values because they provide a compact way of presenting distributions of numerical values and are widely used in the medical domain. The box plots were also more intuitive than the novel speedometer metaphor that was used in $\mathbf{v}_1$–$\mathbf{v}_3$.

The *pT-status* and *pN-status* are both ordinal data, and this required a chart that could show either the number or the proportion of patients. We used a bar chart, which is the obvious choice. It would have been possible to use a pie chart, but for the *pT-status* that would have broken the guideline that there should be no more than six segments [47], and the symmetry of the static dashboard was increased by using the same chart type for both the *pT-status* and the *pN-status*.

The Gleason score is also ordinal but has two dimensions. These could be shown together with a grid-based encoding, which was provided by a heat map. However, during the design iterations we decided that it was better to encode the frequency of each combination of Gleason scores using size rather than color, because that allowed the most common combinations to stand out while reducing the need for color.

The most challenging attributes were the treatment and outcome variables. We used pie charts rather than bar charts because: (a) we wanted those variables to stand out from the other groups of variables (bar charts were the preferred option for the *pT-status* and *pN-status*), and (b) we consider pies to be more appropriate for showing small proportions that should stand out due to their domain importance (e.g., *Meta*, *DOD*, and *CHT*). After investigating three designs of pie chart (see $\mathbf{v}_1$, $\mathbf{v}_3$, and $\mathbf{v}_4$), we chose a design that showed one attribute in each pie chart, trading-off the simplicity of each chart with the number of charts that were required.

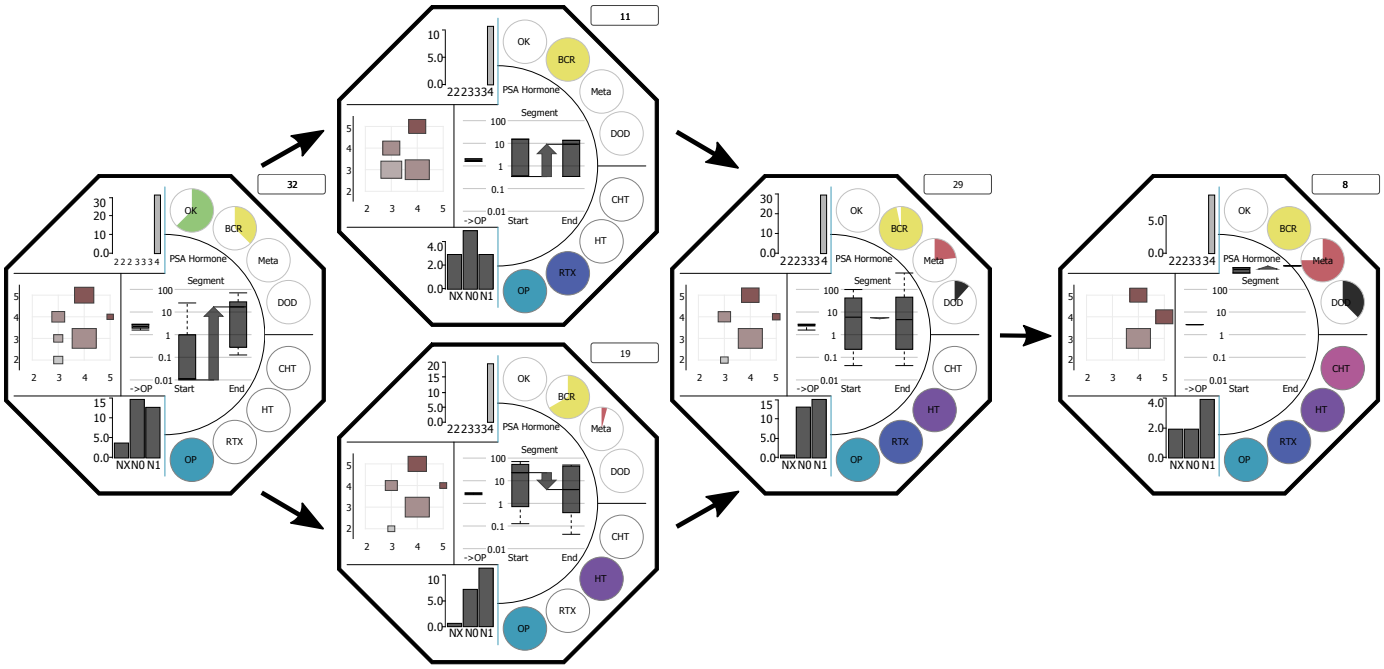

Fig. 3. The history of a cohort of patients who all had a pT-status of pt4. From left to right the static dashboards are ordered by the number of treatments that the patients received (1-4). Two different signatures of treatments exist in the second segment (OP-RTX and OP-HT). Twenty-nine patients have a third segment and eight have a fourth, showing that this cohort contains patients who have the most severe treatment (CHT) and get metastases (Meta) and die (DOD). The colors are the same as shown in Figure 1.

4.6 Structure

The final design ($\mathbf{v}_9$) structured the data by applying principles for *ordering*, *grouping*, *separation*, and *symmetry*. Ordering worked in three ways. The first was for data that was guaranteed to be gathered in a particular order, which was presented in temporal order from left (earliest) to right (latest). An example is the three PSA measurements that were gathered pre-operatively, and at the start and end of each segment. The second was to position the treatment variables in order of severity (*OP*, *RTX*, *HT*, and *CHT*), as defined by medical experts (see $\mathbf{v}_8$ and $\mathbf{v}_9$). The third was to apply a data-driven approach and calculate the order from the sequence in which events typically happened, with an example being the four outcome variables (see $\mathbf{v}_8$ and $\mathbf{v}_9$).

Grouping took advantage of the hierarchical nature of the attributes, and was applied to some extent in all nine iterations to the PSA measurements, and the treatment and outcome variables. However, driven by user feedback, the grouping became more explicit as the designs progressed by adding lines to separate the groups – $\mathbf{v}_1$, $\mathbf{v}_4$ and $\mathbf{v}_5$ relied solely on space to separate the groups, whereas $\mathbf{v}_9$ was the only design that used lines to separate each group from its neighbors.

Symmetry concerned the way in which the attribute groups were arranged within the visualization as a whole, and depended on the fact that there were similarities between some of the attribute groups. Symmetry was achieved by balancing the types of chart that were used on opposite sides of either an horizontal or vertical axis, and is most evident in $\mathbf{v}_{2\text{-}6}$ and $\mathbf{v}_9$.

5 Usage Scenario

This section outlines three usage scenarios for the static dashboards, in line with the groups of required functionalities in Section 3.2. The first scenario (Section 5.1) shows how a single static dashboard may provide an overview of the entire set of segments of the dataset that was described in Section 3.3. The second scenario (Section 5.2) shows how users may identify temporal characteristics of a specific patient cohort. For that, we decided to group the dashboard network by the treatment signature but, of course, other groupings are possible, e.g., with respect to outcome variables. The third scenario (Section 5.3) depicts how dashboard networks can be used for the visual comparison of the histories of two patient cohorts.

5.1 Overview of the Entire Dataset

Overall, the segmentation algorithm described in Section 3.3 created 10,485 segments for the 2,000 patients in the dataset. Despite the amount of aggregation that was needed to reduce all of those segments to information that could be displayed in a single static dashboard (see Figure 1), the visualization preserved the detail that users needed to make a number of observations.

The information for static attributes on the left of the dashboard shows that *pt2c*, *pt3a*, and *pt3b* are the most often occurring *pT-status* values, and the most often observed Gleason scores are *3+4* and *4+3*. The *pN-status N0* is observed in over 6,000 segments. At the upper right the visualization shows roughly one quarter of the patient segments were in a state of *OK*, while about three quarters of the segments indicate *BCR*. The portion of patients suffering from metastases (*Meta*) remains fairly small, which particularly applies to the number of patients that died (*DOD*). According to the dataset, 100% of the patients had *OP*, about two thirds of all cases had radiation treatment (*RTX*), about one quarter received hormone treatment (*HT*), and only a small portion of patients had chemotherapy (*CHT*).

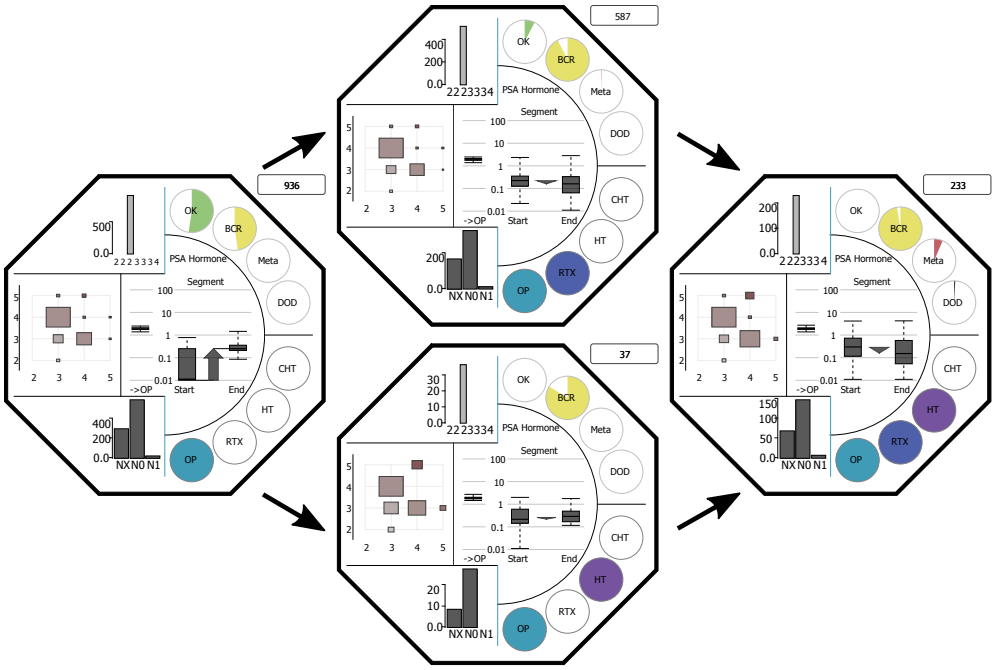

Fig. 4. The history of a cohort of patients who all had a pT-status of pt2c. From left to right the static dashboards are ordered by the number of treatments that the patients received (1-3). Two different signatures of treatments exist in the second segment (OP-RTX and OP-HT). Over nine hundred patients started with segment one and over two hundred have a third segment. However, this cohort contains only few cases with metastases (Meta), virtually no deaths (DOD) and none of the patients received chemotherapy (CHT). The colors are the same as in Figure 1.

5.2 Identify Longitudinal Changes

Figure 3 shows an example dashboard network with five static dashboards that are ordered by the number of treatments that a patient had received (the treatment stage; left to right) and the combination of treatments (the treatment signature; top to bottom). The dashboard network in Figure 3 depicts the history of the cohort of patients who all had a *pT-status* of *pt4*. Users can make observations about longitudinal changes by comparing the dashboards in the network.

In the first stage (left) the only treatment that patients had received is *OP*. The ratio of patients with status *OK* is considerably higher than in Figure 1.By contrast, the outcome variable *BCR* is lower than the overall distribution shown in Figure 1. Moreover, a strong increase in the *PSA* value can be observed. In the second stage, two different treatment signatures exist (*OP* & *RTX* and *OP* & *HT*): patients with *RTX* had lower *PSA* values, but the *PSA* value increased after the treatment in many cases. For patients receiving *HT*, the situation is somewhat opposite, with higher but slightly decreasing *PSA* values. *HT* seems to have been prescribed for patients with more severe biological situations, where in some cases the first *Meta* are diagnosed. Other indicators for the more hazardous biological state of *HT* are the *pN-status*, which contains more *N1*, and the Gleason score *5+4*. Finally, 100% of the patients receiving *RTX* have a *BCR*, hinting at the usual treatment of a biochemical recurrence with radiation therapy. If patients reached the third and fourth stage, the biological condition had become more severe. Many patients had diagnoses of *Meta* and some died *DOD*. The *PSA* values were high and literally shot through the roof in stage four. Chemotherapy was given only in stage four, as a last resort.

5.3 Comparison of two Cohorts of Patients

The dashboard network described in Section 5.2 also allows the comparison between cohorts that are displayed in separate sets of visualizations. The following usage scenario uses Figures 3 and 4 to depict two cohorts with a different *pT-status* (*pt4* and *pt2c*).

One of the most obvious differences between the cohorts is the missing fourth stage in Figure 4, indicating that *pT-status pt2c* does not lead to *CHT*. In stage one both treatment signatures share *OP* as the first form of treatment, which is pre-defined in the dataset. A closer look at the *PSA* trend reveals that the PSA value of the *pt4* cohort increases considerably stronger in the first stage. In stage two, the treatments received by the *pt2c* cohort have less implications on the *PSA* value progression. One hypothesis may be that low PSA values are less responsive to treatments. These observations indicate that *pT-status pt2c* has a better prognosis than *pt4*, which is a hypothesis that can be validated as follows. In stage three only a few *pt2c* patients had *Meta* and few patients died. Similarly, the ratio of the extraprostatic extension of the prostate cancer (*N1*) is substantially higher for *pt4*, while *N0* dominates for *pt2c*. The *PSA* value remains at a low level in stage three for most patients in *pt2c*, particularly when compared with *pt4*. The visual comparison of Gleason scores reveals a dominance of less severe scores for *pt2c* (e.g., *3+4*) when compared with the *pt4* cohort (*4+3*, *4+5*). The overall visual comparison of the two cohorts indicates that *pt2c* is a considerably more pleasant diagnosis than *pt4*.

6 User Evaluation

We present the results of a user evaluation to assess the usefulness of the dashboard networks, using three different user groups: *non-experts*, *visualization experts*, and *medical experts*. These groups span the intended audience for our dashboard networks. Medical experts could benefit from visualization to help understand

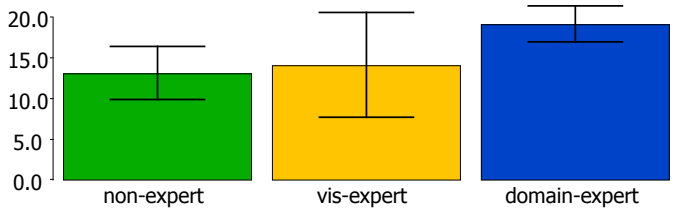

Fig. 5. The number of observations made in $\mathbf{T}_1$ by the three user groups. The standard error (*SE*) is depicted with error bars. Non-experts identified about 13 of all observations, medical experts about 18.

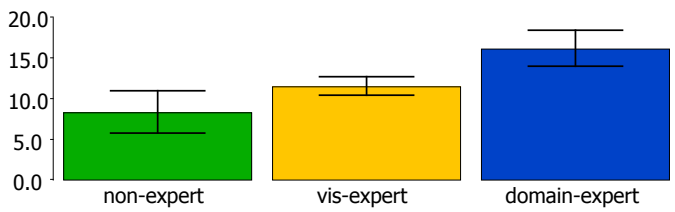

Fig. 6. The number of observations made in $\mathbf{T}_2$ by the three user groups. Error bars show the *SE*. The analytical task is substantially different from $\mathbf{T}_1$, but the pattern between the groups is similar.

variations in patient histories over time and between cohorts. The general public are becoming ever more involved in decision-making about their own treatment, which should be made easier by appropriately designed visualizations, and the same is true for public-patient involvement in health research. Visualization experts lie in-between the other two groups, with the medical knowledge of the general public but with specialist knowledge about the pitfalls and advantages of different styles of visualization and encoding.

6.1 Method

6.1.1 Participants

We recruited 14 participants (2 female, 12 male). All participants had normal vision and did not report color blindness. Their age ranged from 22 to 51 (Median = 32, SD = 8.01). Each participant had at least a diploma/master's degree, five had a PhD degree or higher. Five participants were classified as non-experts because their background was neither in visualization nor prostate cancer. Five participants worked in information visualization research or application, but not in prostate cancer, and are referred to as visualization experts. The remaining four participants were prostate cancer clinicians and are referred to as medical experts.

6.1.2 Procedure

The evaluation took place in a quiet room that contained a participant and always the same interviewer. The evaluation was divided into three parts: introduction (approximately 10 min), familiarization (10 min), and tasks (20 min).

The introduction started with an explanation of a background to prostate cancer (if necessary) and the 15 attributes encoded in the visualization. Then the current state of the art was outlined as described in the introduction section of this paper, and our approach (the visualization of multiple attributes of multiple patient histories) was explained. We showed example images of the dashboard networks to introduce the design and the prostate cancer data in combination with visualization.

The familiarization phase included two questions to ensure that participants understood the data and the static dashboards and networks. First, we used Figure 1 to provide an overview of all segments of the data and to give an impression of the overall distribution of the patient histories (see Section 5.1 for details). The second question was whether or not participants were able to identify the treatment sequences depicted in Figure 3.

In the main part of the evaluation, the participants were asked to conduct two tasks using screenshots (see Figures 1, 3, and 4). We explained that Figure 1 could be used as a legend for the static dashboard and a summary of the data segments provided with the dataset. When performing the tasks, communication with the participant was kept at a minimum, so only general questions of understanding were answered. The tasks were:

- $\mathbf{T}_1$ Identification of characteristics of a patient cohort
- $\mathbf{T}_2$ Visual comparison of two patient cohorts

In $\mathbf{T}_1$ the participant was asked to identify the characteristics and longitudinal changes of a single patient cohort that had a *pT-status pt4* (see Section 5.2 for details). The participant was provided with a static dashboard showing an overview of all of the patients (see Figure 1) and a dashboard network showing the *pt4* cohort (see Figure 3). The treatment signature of the cohort was shown in four stages from a single treatment (always *OP*) on the left up to stage four (four treatments) on the right. At the second stage two different treatment signatures exist and are shown above the other. Some patients received *OP* and *RTX*, while other patients received *OP* and *HT*.

In $\mathbf{T}_2$ the participant was asked to identify differences between the characteristics of the two patient cohorts (see Section 5.3). The two cohorts had a *pT-status pt4* and *pt2c*, respectively, and the participant was provided with Figure 1 (overview), Figure 3 (*pt4* cohort) and Figure 4 (*pt2* cohort). Again, the history of each cohort was presented so that the number of treatments increased from left to right.

6.2 Results

In each task, the participants were asked to describe the characteristics of given patient cohorts, which we refer to as *observations* [48]. If participants made similar observations then they were merged.

The first question in the familiarization phase concerned the overview of all of the patients that is depicted in Figure 1. A table in the supplemental materials provides lists the 12 observations that were made by the participants. The second question was whether or not participants were able to identify the treatment sequences depicted in Figure 3. All 14 participants were able to describe that every patient in the cohort started with *OP* in the first stage, followed by a second treatment in stage two (*OP-RTX* or *OP-HT*). In the third stage the participants had hardly any problems in identifying that *OP*, *RTX*, and *HT* were taken, combining the treatment orders *OP-RTX-HT*, as well as *OP-HT-RTX*. Finally, all participants noticed that *CHT* was always conducted last.

The remainder of this section reports the results for $\mathbf{T}_1$ and $\mathbf{T}_2$, and is divided into three sections. Information about all of the observations that participants made in those tasks is provided in tables in the supplemental materials. First we compare the observations that were made by the three groups of participants. Then we analyze the complexity of the observations, and followed this by analyzing the data attributes that were used. Together, these

Task	Number of attributes	Attribute category	Number of static dashboard segments						
			1	2	3	4	5	6	9
Task 1	1	Biological		2				3	
		PSA				1	2	1	
		Static						4	
		Treatment					1		
	2	PSA	1	4					
		Static				1	1		
		Treatment	1	1	2	1	1		
	3	Static			1				
		Treatment		3	2				
Task 2	2	Biological		2		1			
		PSA		1	1				3
		Static				2			6
		Treatment		2					
	3	Biological		2					
		PSA							1
		Static				1			
		Treatment		2		1			
	4	PSA				1			

Fig. 7. Fine-grained analysis of observation complexities for $\mathbf{T}_1$ and $\mathbf{T}_2$. Colored cells in the grid depict the number of observations the participants had, corresponding with the attribute count, attribute type, and the number of static dashboards (segments) that were used.

results show how each component (chart) of the static dashboards were used, and the extent to which participants were able to integrate many different components from the dashboard networks to make the observations.

6.2.1 Between-groups Comparisons

In $\mathbf{T}_1$ the number of participants who made each observation was significantly correlated for all three pairs of user groups: non-experts vs. visualization experts ($r(31) = 0.63$, $p < .01$), non-experts vs. medical experts ($r(31) = 0.52$, $p < .01$) and visualization experts vs. medical experts ($r(31) = 0.64$, $p < .01$). However, medical experts made more observations (see Figure 5) and showed greater consistency. Thirteen observations were made by all of the medical experts, but only three observations were made by all of the visualization experts and two observations by all of the non-experts. One striking difference was that the medical experts focused more on treatments, with five of the treatments observations being made by all of the medical experts. No treatment observation was made by all members of either the non-experts or visualization experts group. This may be an indication that the mental models of medical experts are targeted towards the treatment of patients, while the other two groups focused on more general observations.

There was a similar pattern of results for $\mathbf{T}_2$. The number of participants who made each observation was significantly correlated for all three pairs of user groups: non-experts vs. visualization experts ($r(24) = 0.73$, $p < 0.01$), non-experts vs. medical experts ($r(24) = 0.46$, $p < 0.05$) and visualization experts vs. medical experts ($r(24) = 0.58$, $p < 0.01$). Medical experts made more observations (see Figure 6) and showed greater consistency, with 15 observations made by all of the medical experts, but only seven observations were made by all of the visualization experts and five observations by all of the non-experts. Three of the treatment observations were made by all of the medical experts, but no treatment observation was made by all members of either the non-experts or visualization experts group.

6.2.2 Complexity of the Observations

Two ways of quantifying the complexity of each observation are the number of attributes and the number of static dashboards that were involved. There was little difference between the three user groups for either complexity measure, so this section combines the observations for all of the participants.

The number of attributes that were involved in each observation ranged from 1–4, the number of static dashboards ranged from 1–9 and, as expected, the former tended to decrease as the latter increased (see Figure 7). The overall complexity of each observation was calculated by multiplying the number of attributes and the number of static dashboards. This overall complexity was in the range 2–27, with the observations generally more complex in $\mathbf{T}_2$ than $\mathbf{T}_1$ because the former involved two cohorts rather than one (9 vs. 5 static dashboards).

The complexity data demonstrates the flexibility of our dashboard networks. Each static dashboard contained 12 charts, meaning that, at the glance of an eye, participants chose flexibly from a total of 60 ($\mathbf{T}_1$) or 108 different charts ($\mathbf{T}_2$) to make a wide variety of observations about longitudinal changes within and differences between patient cohorts.

6.2.3 Attributes

The multitude of attributes included in the static dashboards raises questions about whether all visual encodings of the attributes were useful and participants preferred some attributes over others. To investigate this, we mapped each observation to its attributes.

The mapping for $\mathbf{T}_1$ is presented in Figure 8. Fourteen attributes are involved, and the most popular attributes are *PSA Trend (PSA T.)*, *RTX*, *HT*, and *CHT*. The focus of medical experts on treatments is clearly shown. By contrast, outcome variables (*PSA Trend*, *PSA Value*, *Meta*, and *DOD*) were identified by all three groups almost equally. This may be due to the fact that non-experts and visualization experts focused on the well-being status of patients rather than patient treatment.

The mapping for $\mathbf{T}_2$ (comparing two patient cohorts) is presented in Figure 9 and shows a different distribution of observations. Overall eleven attributes were taken into consideration. The most common attributes are *pN-status (pN-St.)*, *PSA Value (PSA V.)*, and *PSA Trend (PSA T.)*. At least two reasons may have an influence on the resulting distribution. First, the comparison task emphasizes attributes showing differences between the *pt4* and the *pt2c* cohort, and Figures 3 and 4 clearly reveal major differences between the *Gleason* score, *PSA* values and trends, and the *pN-status*. The second reason may be that the usage scenario of $\mathbf{T}_2$ is based on treatments that were received. The treatment signatures of the two cohorts were similar, although the *pt2c* cohort had to undergo fewer treatments and no *CHT* at all.

6.3 Data Quality Observations

Using the dashboard networks, participants identified two data quality problems with the PSA attributes. These problems show how our dashboard networks may be used to verify the quality and integrity of data.

First, some of the medical experts queried the *pre-OP PSA* because the value was rather unexpected (the last PSA value before OP has an influence on patients' development in the follow-up phase). On investigation, we found that the data model contained an error, meaning that *pre-OP PSA* was incorrectly shown to be constant across the dashboard network.

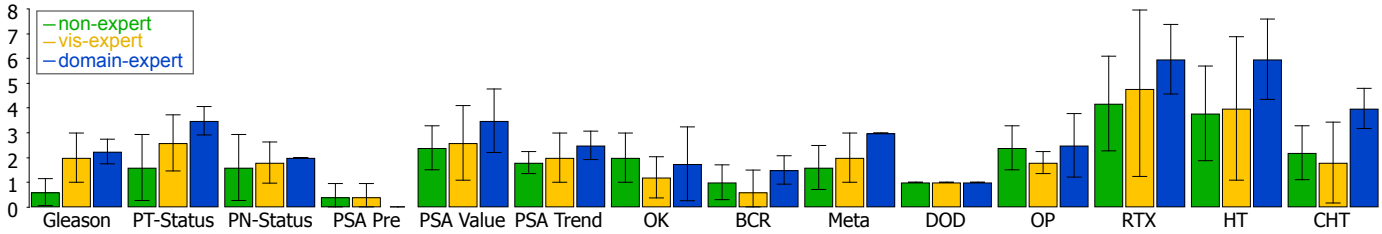

Fig. 8. The mean number of times participants made an observation that was about each attribute in $\mathbf{T}_1$ (characteristics of the patient cohort). Error bars show the *SE*. Abbreviation: PSA PRE is *pre-OP PSA*. Medical experts emphasized treatment attributes while outcome variables were identified equally. Overall, *RTX* and *HT* were most interesting to solve $\mathbf{T}_1$.

The second problem concerned some unexpected zero values of the *PSA* values for the *pt2c* cohort (see Figure 4), which in two of the static dashboards caused the whiskers of the boxplot to spread down to 0.0. Two visualization experts identified this issue, leading to discussions about the way the value domain should be represented.

6.4 Discussion

The overall success of the design is demonstrated by the fact that all three groups of participants (non-expert, visualization experts and medical experts) made a sizable number of observations in each task, with medical experts making the most. At the glance of an eye, participants were able to pick and mix information from 2 – 27 charts out of a much larger number of charts in the dashboard networks (60 and 108 charts in $\mathbf{T}_1$ and $\mathbf{T}_2$, respectively) to make each observation. Every single attribute was used in some observations and, of course, the frequency with which each attribute was used was affected by the application domain.

The evaluation had a number of limitations. The most important one was that the data was more orderly and constrained than in electronic health records (EHRs) as a whole, because there were only a small number of treatments and outcomes (four of each, compared with tens of thousands of codes in disease classifications such as ICD-10) and the data was curated so its quality was higher than is often found in EHRs. Other limitations were the use of a fairly small number of participants, and the fact that all of the participants were familiarized with the dashboard networks at the beginning of the evaluation. In a real-world setting, typically it would not be possible to spend time familiarizing non-experts (e.g., patients) prior to a doctor/patient consultation.

Although the design was specific to a particular medical domain, there are aspects that provide guidance for the application of the design to other domains. Some of that guidance results in three design principles that help to provide structure for static dashboards: (1) divide charts into groups according to their attributes, (2) position the charts so that they are ordered according to time, and (3) explicitly separate each group of charts. A fourth principle is to arrange charts symmetrically, where possible, because the improved aesthetics [49] is likely to make it easier for users to compare static dashboards longitudinally or between-cohorts.

Encoding multiple attributes causes conflicts in the choice of visual channels, so we considered several design alternatives for each attribute encoding. One example can be seen in design $\mathbf{v}_7$ (Figure 2) where the outcome variables are encoded with a barchart instead of a pie chart. As a result, the symmetry between outcome variables (top) and treatments (bottom) was lost and users started to seek connections between the outcome variables (right) and the *pT-status* (left), which was not intended. Another example is the limited use of color, which allowed more flexibility for dealing with semantic overlap between the meaning of different attributes (e.g., *Meta* and *CHT* both indicate severe patient conditions and hence the use of a reddish color). In this regard, we refer to a quote from David Travis "When correctly used, the benefits of color are unrivaled ... However, when incorrectly used, color has the potential to make a system unusable" [50].

The design also had some limitations, which we briefly summarize here. In the usage scenarios used for the evaluation, we aggregated the data for patients who received the same treatment at a given stage (e.g., stage 3 in Figure 3), but some participants said that it may be useful to separate patients who had received different sequences of treatments (e.g., *OP-RTX-HT* vs. *OP-HT-RTX*). The design could be improved for attributes that remain constant, so users do not have to compare static dashboards to deduce that an attribute has not changed. The segmentation process was designed to differentiate between changes in the attributes, but some types of treatment (e.g., hormone therapy) take place over an extended period of time. A color legend should have been provided for users who were not familiar with the chosen domain-specific norms. Informal feedback indicated that the presentation of the PSA attributes could be improved but the best solution remains an open question. It could be argued that the PSA trend is redundant, because that is implicit from the PSA start and PSA end boxplots. However, many participants expressed observations in terms of the trend, rather than absolute PSA values, and one participant suggested that the trend should be discretized to three directions (constant, up, and down) which would allow the current trend arrow to be made more salient.

Finally, the scalability of the design to more diverse patient cohorts remains an open question. As more treatment pathways are introduced (e.g., imagine four treatments that could occur in any order, rather than the two orders of the present research) the complexity of the dashboard networks will increase. This will increase the cognitive effort of making comparisons and at some point, as with all visualizations, the design will break down.

7 Conclusion

This paper describes a design study that developed a novel method for visualizing patients' medical histories so that longitudinal and cross-cohort patterns could be understood. Our solution presented those histories using a three-tier visualization structure that comprises: (1) charts that presented each attribute in the patient histories, (2) static dashboards that contained the charts for given temporal segments of the histories, and (3) a dashboard network that showed all of the segments for a given patient cohort.

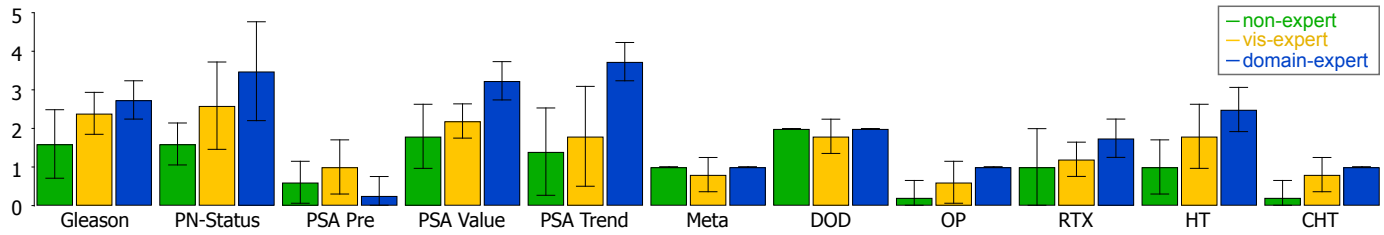

Fig. 9. The mean number of times participants made an observation that was about each attribute in $\mathbf{T}_2$ (comparison of two patient cohorts). Participants in all three groups emphasized biological attributes and early indicators for the disease to answer $\mathbf{T}_2$. Error bars show the *SE*.

Central to the dashboard network is our data segmentation method, which preserves the order in which events occur, rather than the events' actual time. This allows the static dashboards to reveal the proportion of a population (e.g., patients) with each attribute, and the dashboard network to show longitudinal and between-cohort patterns.

The dashboard network design evolved during nine iterations, during which we interviewed medical experts, tested design alternatives, and held critical discussions with visualization experts. The success of the design is shown by the large number of observations that were made by users who ranged from non-experts to medical experts, and the diversity of those observations in terms of both the number of attributes and the number of history segments that were involved.

There are a number of future avenues for our research. One is to investigate the ability of dashboard networks to handle missing or noisy data. A second one is to understand the extent to which the networks scale in terms of the number of attributes, static dashboards and dashboard networks that are possible to be displayed while still allowing users to easily make diverse observations. Finally, the effectiveness of dashboard networks should be investigated in other application domains.

Acknowledgments

We would like to thank our collaboration partners at the Martiniklinik at UKE in Hamburg for their time, advice, and all the fruitful discussions. We also thank the 14 participants of the user evaluation. This research was supported by an award to R. A. Ruddle by the Alexander von Humboldt Foundation.

References

[1] B. Shneiderman, C. Plaisant, and B. W. Hesse, "Improving healthcare with interactive visualization," *IEEE Computer*, vol. 46, no. 5, pp. 58–66, 2013.

[2] W. Raghupathi and V. Raghupathi, "Big data analytics in healthcare: promise and potential," *Health Information Science and Systems*, vol. 2, no. 1, p. 3, 2014.

[3] A. Rind, T. D. Wang, W. Aigner, S. Miksch, K. Wongsuphasawat, C. Plaisant, and B. Shneiderman, "Interactive information visualization to explore and query electronic health records," *Foundations and Trends in Human-Computer Interaction*, vol. 5, no. 3, pp. 207–298, 2013.

[4] T. Gschwandtner, K. Kaiser, and S. Miksch, "Information requisition is the core of guideline-based medical care: which information is needed for whom?" *Journal of Evaluation In Clinical Practice*, vol. 17, pp. 713–721, 2011.

[5] F. Du, B. Shneiderman, C. Plaisant, S. Malik, and A. Perer, "Coping with volume and variety in temporal event sequences: Strategies for sharpening analytic focus," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 23, no. 6, pp. 1636–1649, 2017.

[6] S. M. Powsner and E. R. Tufte, "Graphical summary of patient status," *The Lancet*, vol. 334, pp. 386–389, 1994.

[7] S. Powsner and E. Tufte, "Summarizing clinical psychiatric data," *Psychiatric Services*, vol. 48, pp. 1458–1461, 1997.

[8] C. Plaisant, B. Milash, A. Rose, S. Widoff, and B. Shneiderman, "Lifelines: Visualizing personal histories," in *SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '96. New York, NY, USA: ACM, 1996, pp. 221–227.

[9] R. Bade, S. Schlechtweg, and S. Miksch, "Connecting time-oriented data and information to a coherent interactive visualization," in *SIGCHI Conference on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2004, pp. 105–112.

[10] J. Bernard, D. Sessler, T. May, T. Schlomm, D. Pehrke, and J. Kohlhammer, "A visual-interactive system for prostate cancer cohort analysis," *IEEE Computer Graphics and Applications (CG&A)*, vol. 35, no. 3, pp. 44–55, 2015.

[11] J. Bernard, D. Sessler, A. Bannach, T. May, and J. Kohlhammer, "A visual active learning system for the assessment of patient well-being in prostate cancer research," in *Workshop on Visual Analytics in Healthcare (VAHC '15)*. New York, NY, USA: ACM, 2015, pp. 1:1–1:8.

[12] M. Gleicher, D. Albers, R. Walker, I. Jusufi, C. D. Hansen, and J. C. Roberts, "Visual comparison for information visualization," *Information Visualization*, vol. 10, no. 4, pp. 289–309, 2011.

[13] T. D. Wang, C. Plaisant, B. Shneiderman, N. Spring, D. Roseman, G. Marchand, V. Mukherjee, and M. Smith, "Temporal summaries: Supporting temporal categorical searching, aggregation and comparison," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 15, no. 6, pp. 1049–1056, 2009.

[14] 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*, 2011.

[15] A. Perer and D. Gotz, "Data-driven exploration of care plans for patients," in *CHI '13 Extended Abstracts on Human Factors in Computing Systems*, ser. CHI EA '13. New York, NY, USA: ACM, 2013, pp. 439–444.

[16] D. Gotz and H. Stavropoulos, "Decisionflow: Visual analytics for high-dimensional temporal event sequence data," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 20, no. 12, pp. 1783–1792, 2014.

[17] 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 *SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. New York, NY, USA: ACM, 2011, pp. 1747–1756.

[18] E. A. Carter, R. S. Burd, M. Monroe, C. Plaisant, and B. Shneiderman, "Using eventflow to analyze task performance during trauma resuscitation," in *Workshop on Interactive Systems in Healthcare*, 2013.

[19] M. Monroe, T. E. Meyer, C. Plaisant, R. Lan, K. Wongsuphasawat, P. Col, T. S. Coster, S. Gold, J. Millstein, and B. Shneiderman, "Visualizing patterns of drug prescriptions with eventflow: A pilot study of asthma medications in the military health system," in *Technical report, DTIC Document*, 2013.

[20] A. Lex, M. Streit, H.-J. Schulz, C. Partl, D. Schmalstieg, P. J. Park, and N. Gehlenborg, "Stratomex: Visual analysis of large-scale heterogeneous genomics data for cancer subtype characterization," *Computer Graphics Forum (CGF)*, vol. 31, no. 3, pp. 1175–1184, 2012.

[21] M. H. Loorak, C. Perin, N. Kamal, M. Hill, and S. Carpendale, "Timespan: Using visualization to explore temporal multi-dimensional data of stroke patients," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 22, no. 1, pp. 409–418, 2016.

[22] J. Krause, A. Perer, and H. Stavropoulos, "Supporting iterative cohort construction with visual temporal queries," *Visualization and Computer Graphics, IEEE Transactions on*, 2015.

[23] S. Malik, F. Du, M. Monroe, E. Onukwugha, C. Plaisant, and B. Shneiderman, "Cohort comparison of event sequences with balanced integra-

tion of visual analytics and statistics," in *Intelligent User Interfaces (IUI)*. New York, NY, USA: ACM, 2015, pp. 38–49.
[24] F. Du, C. Plaisant, N. Spring, and B. Shneiderman, "Eventaction: Visual analytics for temporal event sequence recommendation," in *2016 IEEE Conference on Visual Analytics Science and Technology (VAST)*, 2016, pp. 61–70.
[25] A. Inselberg and B. Dimsdale, "Parallel coordinates: A tool for visualizing multi-dimensional geometry," in *Conference on Visualization (VIS)*. Los Alamitos, CA, USA: IEEE Computer Society, 1990, pp. 361–378.
[26] P. Hoffman, G. Grinstein, K. Marx, I. Grosse, and E. Stanley, "DNA Visual and Analytic Data Mining," in *Conference on Visualization (VIS)*. Los Alamitos, CA, USA: IEEE Computer Society, 1997, pp. 437–ff.
[27] D. A. Keim, "Designing pixel-oriented visualization techniques: Theory and applications," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 6, no. 1, pp. 59–78, 2000.
[28] Cleveland, W. S. and McGill, R., "The Many Faces of a Scatterplot," *Journal of the American Statistical Association*, vol. 79, no. 388, pp. 807–822, 1984.
[29] J. Fuchs, "Glyph Design for Temporal and Multi-Dimensional Data: Design Considerations and Evaluation," Ph.D. dissertation, Universitt Konstanz, Konstanz, 2015.
[30] R. Borgo, J. Kehrer, D. H. Chung, E. Maguire, R. S. Laramee, H. Hauser, M. Ward, and M. Chen, "Glyph-based visualization: Foundations, design guidelines, techniques and applications," in *Eurographics State of the Art Reports*, ser. EG STARs. Eurographics Association, 2013, pp. 39–63.
[31] M. H. Loorak, C. Perin, C. Collins, and S. Carpendale, "Exploring the possibilities of embedding heterogeneous data attributes in familiar visualizations," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 23, no. 1, pp. 581–590, 2017.
[32] J. Bertin, *Semiology of Graphics*. University of Wisconsin Press, 1983.
[33] E. R. Tufte, *The Visual Display of Quantitative Information*. Cheshire, CT, USA: Graphics Press, 1986.
[34] E. Tufte, *Envisioning Information*. Graphics Press, 1990.
[35] C. Perin, P. Dragicevic, and J. D. Fekete, "Revisiting Bertin Matrices: New Interactions for Crafting Tabular Visualizations," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 20, no. 12, pp. 2082–2091, 2014.
[36] E. Maguire, P. Rocca-Serra, S.-A. Sansone, J. Davies, and M. Chen, "Taxonomy-based glyph design—with a case study on visualizing workflows of biological experiments," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 18, no. 12, pp. 2603–2612, 2012.
[37] W. Horn, C. Popow, and L. Unterasinger, "Metaphor graphics to visualize icu data over time," in *Intelligent Data Analysis in Medicine and Pharmacology (IDAMAP), Workshop Notes of the ECAI-98 Workshop*, 1998, pp. 76–81.
[38] N. Wilhelm, A. Vgele, R. Zsoldos, T. Licka, B. Krger, and J. Bernard, "FuryExplorer: visual-interactive exploration of horse motion capture data," in *SPIE, Visualization and Data Analysis (VDA)*, 2015, pp. 93 970F–93 970F–15.
[39] N. Cao, D. Gotz, J. Sun, and H. Qu, "Dicon: Interactive visual analysis of multidimensional clusters," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 17, no. 12, pp. 2581–2590, 2011.
[40] T. Munzner, "A nested model for visualization design and validation," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 15, no. 6, pp. 921–928, 2009.
[41] M. Sedlmair, M. Meyer, and T. Munzner, "Design study methodology: Reflections from the trenches and the stacks," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 18, no. 12, pp. 2431–2440, 2012.
[42] T. Gschwandtner, H. Schuman, J. Bernard, T. May, M. Bögl, S. Miksch, J. Kohlhammer, M. Röhlig, and B. Alsallakh, "Enhancing time series segmentation and labeling through the knowledge generation model," in *Eurographics Conference on Visualization (EuroVis), Poster*. Eurographics Association, 2015, p. 3.
[43] J. Bernard, M. Steiger, S. Widmer, H. Lücke-Tieke, T. May, and J. Kohlhammer, "Visual-interactive Exploration of Interesting Multivariate Relations in Mixed Research Data Sets," *Computer Graphics Forum (CGF)*, vol. 33, no. 3, pp. 291–300, 2014.
[44] H. Shimazaki and S. Shinomoto, "A method for selecting the bin size of a time histogram," *Neural Comput.*, vol. 19, no. 6, pp. 1503–1527, 2007.
[45] T. Mühlbacher and H. Piringer, "A partition-based framework for building and validating regression models," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 19, no. 12, pp. 1962–1971, 2013.
[46] T. Munzner, *Visualization Analysis and Design*, ser. A.K. Peters visualization series. A K Peters, 2014. [Online]. Available: http://www.cs.ubc.ca/ tmm/vadbook/
[47] M. Hardin, D. Hom, R. Perez, and L. Williams, "Which chart or graph is right for you?" Tableau Software, 2016.
[48] P. Saraiya, C. North, and K. Duca, "An insight-based methodology for evaluating bioinformatics visualizations," *IEEE Transactions on Visualization and Computer Graphics (TVCG)*, vol. 11, no. 4, pp. 443–456, 2005.
[49] H. C. Purchase, "Effective information visualisation: a study of graph drawing aesthetics and algorithms," *Interacting with computers*, vol. 13, no. 2, pp. 147–162, 2000.
[50] D. Travis, *Effective color displays : theory and practice*, ser. Computers and people series. London: Academic Press, 1991. [Online]. Available: http://opac.inria.fr/record=b1089201

Jürgen Bernard is Post-doc at TU Darmstadt, Germany. He was with Fraunhofer IGD when he received is PhD degree in 2015. His research is in information visualization and visual analytics with an emphasis on medical health care applications. His data-centered focus is on time-oriented data and high-dimensional data. Many of his solutions combine techniques from cluster analysis, dimensionality reduction, similarity search, or other machine learning techniques.

David Sessler David Sessler is a scientific assistant at the Fraunhofer Institute for Computer Graphics Research and a Master degree student in Visual Computing at the TU Darmstadt. Sessler has a Bachelor of Science degree in Computer Science from the TU Darmstadt. His research is in combining machine learning and visual analytics, e.g., visual-interactive similarity search for complex data objects and time series, or visualizations for medical data.

Jörn Kohlhammer is the head of the Competence Center for Information Visualization and Visual Analytics at the Fraunhofer Institute for Computer Graphics Research. His research interests include decision-centered information visualization and visual analytics. Kohlhammer has a PhD in computer science from the Technische Universitt Darmstadt.

Roy Ruddle is Professor of Computing at the University of Leeds. He has a multidisciplinary background, combining research and development in the software industry with a PhD in psychology. He conducts basic and applied research at the interface of computer graphics and human-computer interaction. His current research focuses on ultra-high-definition displays and visual analytic tools.