Visual analytics for evaluating clinical pathways

Please check the document version of this publication:

• A submitted manuscript is the version of the article upon submission and before peer-review. There can be important differences between the submitted version and the official published version of record. People interested in the research are advised to contact the author for the final version of the publication, or visit the DOI to the publisher's website.
• The final author version and the galley proof are versions of the publication after peer review.
• The final published version features the final layout of the paper including the volume, issue and page numbers.

Link to publication

Visual Analytics for Evaluating Clinical Pathways

Humberto S. Garcia Caballero*   Alberto Corvò†   Prabhakar M. Dixit‡   Michel A. Westenberg§
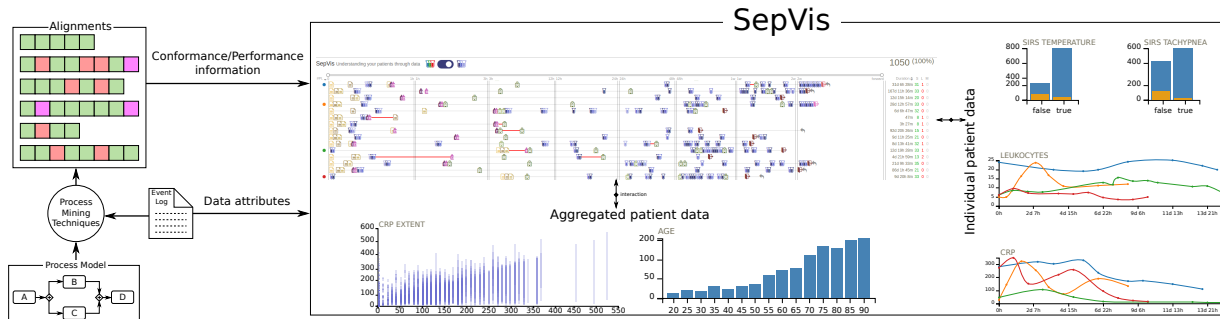
Eindhoven University of Technology

Figure 1: Overview of SepVis. A visual analytics tool that takes performance and conformance information from alignments. It provides different views to explore clinical pathways as well as patient-related data. The data attributes are extracted from an event log. A process model describing a guideline in a hospital setting in conjunction with the event log are used to compute the alignments by an algorithm.

ABSTRACT

Digital platforms in healthcare institutions enable tracking and recording of patient care pathways. Besides the Electronic Health Records (EHRs), the event logs from Hospital Information Systems (HIS) are a very efficient source of information, from both operational and clinical point of view. Process mining allows comparison of a patient care pathway with the event log(s) from HIS, to understand how well the reality as depicted in the event log fits the expectation as modeled using a care pathway. In this paper, we present SepVis, a visual analytics tool which aims to fill the gap in current process-centric applications by looking at patients' pathways from a clinical point of view. We demonstrate the utility of SepVis in selected use cases derived by the guidelines in the management of sepsis patients.

Keywords: Visual Analytics, Event-logs, Process Mining, Conformance-checking, Sepsis

Index Terms: Computer Graphics [I.3.8]: Applications—

1 INTRODUCTION

A visit of a patient to a hospital results in a certain sequence of activities that are performed by specific resources. Some activities, such as administrative tasks, could be valid for most of the incoming patients. However, some other activities are specific to a cohort of patients, depending on factors such as the type of disease, diagnosis treatment plan, etc. Generally, a hospital has a well-defined care pathway in place depending on the type of incoming patient. These patients are the primary actors (cases) of the care pathways, which inherently consider the essential protocols and/or guidelines.

With the advent of digitization, most of the steps performed, such as registration of ward activities by clinical personnel, treat-

*e-mail: h.s.garcia.caballero@tue.nl

†e-mail: a.corvo@tue.nl

‡e-mail: p.m.dixit@tue.nl

§e-mail: m.a.westenberg@tue.nl

ment plans and results, logistics and medical decisions, are well documented in the Hospital Information Systems (HIS) [18]. For a limited number of patients in a hospital, the data from the information systems can be retrospectively explored to analyze clinical decisions, deviations from the care pathways, etc. However, the task of analyzing the data from the HIS in a care pathway setting can be extremely challenging if the number of patients is large. This is especially the case for a single type of ward and/or disease because patients may have different profiles and undergo different *versions* of the care pathway. This adds new dimensions, and hence complexity, to the data and makes the orchestration and integration of data challenging. However, this additional information also enriches the value of tracked activities in the patients' flows, by providing additional attribute information.

One key enabler to analysis of huge care pathway specific data is Process Mining [20]. Traditionally process mining is divided into process discovery and conformance analysis. Process discovery, as the name suggests, discovers a process (care pathway) model from any event log, which is typically derived from HIS or EHRs in a healthcare setting. Conformance analysis uses a process model and projects an event log on the process model [21]. Conformance analysis provides insights into which part of the process fits the data, and which part does not, thereby giving insights into deviations that occur compared to the expected behavior. Furthermore, process mining output can also shed light on performance-related information, such as throughput time of a patient in a care pathway setting, bottlenecks in the process, slow paths in the process, etc. This information could be used to assess the performances in hospital [17] settings and/or to cut down costs and delays. Root cause analysis could be performed on the computed process information to investigate the reasons for deviations and reasons behind different decisions.

In the analysis of event logs from HIS, most of the current approaches do not consider the *care pathway* dimension. Instead, the focus is usually on aligning sequences of variants guided by sequence mining algorithms. However, such techniques often bypass the importance of care processes and the clinical protocols and guidelines that are modeled in such pathways. On the other hand, process mining centered approaches take into account these models, but the visualization focus is entirely addressed from a *process* perspective. That is, the outcome is consolidated and projected on a
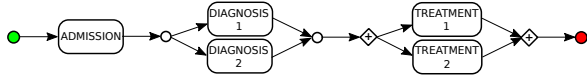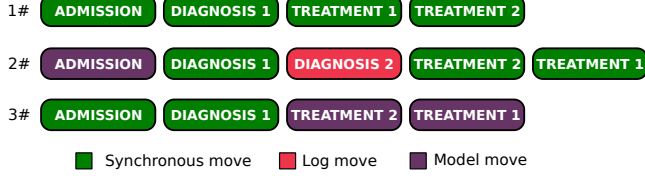
Process Model

Possible alignments

Figure 2: A simple process model in a hospital setting and a set of possible alignments. The first step in the process is Admission, followed by either one of Diagnosis 1 or Diagnosis 2. Diagnosis is followed by both Treatment 1 and Treatment 2, which can occur in any order.

process model. This makes it difficult to understand and visualize patterns or cohorts of patients that may exist. By focusing entirely on the process dimension, other relevant patient's attributes which may be important are also not visualized.

In this paper, we present our work in progress on SepVis, a visualization tool that aims to extend the standard visual analytics tools in the healthcare domain with support of process mining techniques. Instead of using traditional sequence alignment techniques, which do not consider the important *care pathway* dimension, we use the results from process mining techniques in order to lay out and visualize the activities of patient flows in accordance with the *clinical model*. Furthermore, instead of using a process model, by visualizing the activities directly, a transparent and direct *view on data* is provided for the user to interactively explore and gain valuable insights into the data.

2 BACKGROUND

This section provides a brief description of the underlying components which are used in our approach. We begin by describing the so-called *event logs*, retrieved from HIS. Next, we briefly introduce process models and alignment based conformance analysis concepts, which are widely researched in the field of process mining [20].

2.1 Event Log

The HIS, such as the Electronic Health Records systems, can be used to extract event logs. Events form the basic building blocks of an event log. Each event is associated with a particular case, which is identified by a case ID, which typically is a patient ID in a healthcare setting. An event represents occurrence of an activity in a particular case. Each event has a time-stamp associated with it, which denotes its time of occurrence. Furthermore, each event may contain additional event specific data attributes, for example the resource which performed the particular activity represented by the event. Similarly, each case may have case level data attributes, such as the gender of the patient, age of the patient, etc. The event log is thus a collection of sequences of events that represent individual cases. Event logs extracted from the HIS can be used to perform process mining analysis.

2.2 Process Models and Alignment based Conformance

Process models are used to model the desired *flow* of steps followed by patients within a healthcare setting. Fig. 2 shows a simple example of a process model. The guidelines and protocols from a hospital can be modeled using such process models. It should be noted that alongside sequences, most of the process modeling notations allow for rich behavioral aspects, such as choice in between multiple ac-
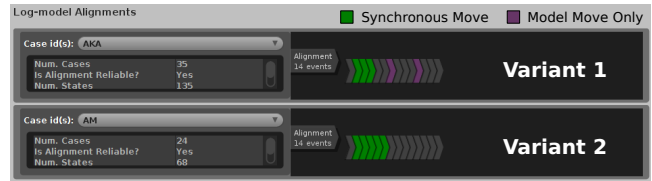

Figure 3: Visualization of alignments in ProM. Patients are grouped by variants. The outputs from conformance checking are visible on the variants as colored arrow heads. However, this view does not include information regarding duration of individual traces, time between activities and it does not allow the user to explore the process by activities' attributes filtering.

tivities, concurrent execution of activities, repetition of a particular fragment in a model, etc.

In our approach, we use the alignment-based conformance analysis technique for getting accurate conformance and performance information pertaining to the process. This technique [21] finds a *path* from start to end in the process model that best describes a case from the event log by using a cost function. The operation of this algorithm is not covered in our paper. It also provides insights into how well an event log fits a particular process model. For instance to investigate how and where the process deviates from the protocols/guidelines in place. Next to this, alignments can also be used to investigate the possible bottlenecks in the process. For each event in the event log, the following three possibilities can arise (see alignments in Fig. 2):

Synchronous move: Occurrence of an event in a trace that is allowed by the current state of the model. The state of the process is given by the execution of previous events in a trace.

Move on log: Occurrence of an event in the trace that is not allowed by the current state of the process. The second trace in Fig. 2 contains a log move because *diagnosis 2* happens in the trace although it is not allowed by the model (only *diagnosis 1* or *diagnosis 2* can happen, but not both.

Move on model: The current state of the process cannot continue with the event that is being treated in the trace. The second trace in Fig. 2 starts with a model move since the model requires to begin with *admission* activity. However, this activity is missing in that trace. Similarly, the third trace contains two model moves at the end since both *treatment 1* and *treatment 2* are mandatory (notice that they can happen in any order).

3 RELATED WORK

In recent years, many applications have been introduced within the visual analytics community for the exploration of temporal event sequences and healthcare data. We distinguish two categories in this domain: sequence alignment based techniques and process analytic based techniques. In this section, we highlight the relevant contributions from the literature belonging to each category.

3.1 Sequence alignment based techniques

There are many techniques in the literature that focus on visualizing the alignment of sequences to represent temporal event data. The initial work in this category provided tools showcasing a representative description of a single patient record [5, 16]. However, the increasing size of HIS records has boosted the development of tools with a focus on exploring, querying and visualizing large population data. Recent advances in this category are applications such as Life-Flow [25] and OutFlow [24], dedicated to the exploration of event sequences in EHRs. EventFlow [13] primarily focuses on medical data, but it has been applied across many domains, e.g. log-analysis,

cyber-security, sport analytics, learning analytics and incident management [12]. In recent work, other visualization approaches for representing pathways within a hospital has been investigated. For instance, a tracking graph construction [23] have been used to describe patient progression over time in an IC unit. Another enabler in finding patterns in longitudinal clinical data is given by interaction which is effectively demonstrated in Borland et al. [2].

However, all of these applications rely entirely on the data from EHRs. These techniques give a powerful view of the reality, used to visualize sequences using algorithms such as sequence clustering. However, contrary to our approach, these techniques do not take process-centric (care-pathway) information into account. Hence, deviations from care-pathway models and clinical protocols cannot be analyzed.

3.2 Process analytic based techniques

Conformance analysis is a well researched field in the area of process mining, which extracts valuable information from event-logs based on certain process models. Among the commercial tools, Celonis [4] provides visual interaction with process centric data and enables the user to define filters according to the available data in the dataset. The process mining framework ProM [22] contains various plug-ins that provide various interesting visualizations of the results from conformance analysis. Fig. 3 provides a snapshot of one such visualization from ProM. From a visualization perspective, many of the current process analytic techniques project the conformance and performance results directly on a process model. There are a couple of issues in such a visualization practice. First, the modeling notation used for representing the process model could be non-trivial. Hence, for a user, e.g. a doctor, the projected information on the process model might be difficult to comprehend. Second, visual analytics features to investigate and make use of other attributes are often missing. In our previous work [7], which is implemented as a plugin in ProM, we combined conformance and performance analysis with interactive visualization of the process model by explicitly making use of data attributes. However, in SepVis, we aim to make a step forward in the support of visual analytics techniques, by extending the current visualization from ProM for conformance and analysis. Instead of interacting with the process models, which could be difficult to understand to the end user, the focus in SepVis is on providing a better user experience for the exploration of variants, traces, and deviations using information from the conformance analysis techniques, thereby taking into account protocols and guidelines that form a part of the care pathway. In essence, SepVis combines the visualization of sequence variants with the capabilities of process mining.

4 DATASET AND TASKS

For the purposes of this work, we use a dataset obtained by an Enterprise Resource Planning (ERP) system in the emergency department of a hospital in The Netherlands [11]. Patients suspected of a sepsis were tracked from the emergency room, where the doctors diagnose the condition. According to the severity of the disease, indicated by symptoms such as infections, body temperature, respiratory rate, heart rate and diffuse psyco-physic pain, patients are administered antibiotics and directed to normal care (NC) units or intensive care (IC) units [10].

Protocol procedures for sepsis treatment are defined [6], but depending on resources and availability of staff and wards, the care pathways can still vary in terms of scheduling and resources.

The dataset comprises 1050 cases described by 15000 events that were recorded for 16 different activities. In addition, 39 data attributes were included to enrich patients pathways. These attributes are divided in two groups: lab tests and information from checklists filled out by nurses. Among the lab tests, we have C-Reactive Protein (CRP), White Blood Cells (Leukocytes) and Lactic Acid

tests. These tests are used by doctors in order to monitor the status of sepsis patients. An interesting fact regarding these tests is that they can occur many times in a single case, and hence our tool has to cope with it. Besides these attributes, we also have attributes indicating the age of the patient, and boolean attributes signaling conditions of the patient. These clinical attributes denote whether the patient matched the Systemic Inflamatory Response Syndrome (SIRS) criteria [1] and organ malfunction, emo-dynamics and inflammatory parameters, which are used in the *sepsis screening triage*. Most of these attributes were recorded at the emergency room entrance. Next, they are used by doctors to plan the specific care pathway.

The activities contained in the dataset belong to different categories. Some activities are related to blood tests, others concern admission to NC or IC units within the hospital and others relate to administrative tasks etc. Interesting records regarding administration of *intravenous liquids* and *intravenous antibiotics* are also available.

A set of questions was obtained by interviewing clinicians responsible for the emergency department [11]. These questions arise when the kind of data described above is available on the system. Next, we defined a set of requirements which we used for the implementation and generalization of our tool. Each task is related to specific variables from the process-point of view:

T1 Patients' *activities* are strongly related to patients' clinical conditions *(attributes)*. Users want to be able to see processes from a clinical point of view.

T2 Time between activities hints to the observance of protocols or not. Delays in relation to guidelines indicate the *performance* of the clinical department.

T3 Evidence for changes in care pathways leads to understanding reasons for mistakes, different clinical decisions or inattentive activity recording. *Deviations* need to be visible and clear.

In **T1**, the clinical staff intends to get an overview of recorded cases from the activity perspective. These can be the admissions to a specific ward, a sequence of activities, medications, or laboratory tests. The link between clinical conditions and activities is highly important to understand decisions within the single care-pathway, as for instance critical patients who are expected to be treated promptly according to clinical protocols.

Task **T2** is performed with the intention to look at performances in wards and execution time between activities. In clinical settings, time represents a key aspect and the insight into delays due to unavailability of resources is decisive in management analysis of life-threatening diseases.

Lastly, **T3** is important because clinical procedures in the hospital are susceptible to doctors' decisions. Groups of patients can deviate from the expected behavior because of doctor decisions based on demographic information or clinical conditions (e.g., age, blood tests) that differ from the model but guaranteed success or seemed reasonable at that point in time.

5 DESIGN DECISIONS

Our system SepVis was developed according to the tasks presented in Sect. 4. In this section we discuss our design decisions.

5.1 Icon design

We identified five main categories of activities within our dataset, which we represent by icons. The categories are: administration, tests, ward admissions, treatments, and release/return actions. The shapes of the icons match the meaning of these categories. For instance, the icon used to display tests is a test tube, the icon showing treatments is a jar, and so on. This way, users can easily identify groups of activities, and have an idea what is happening to the patient.
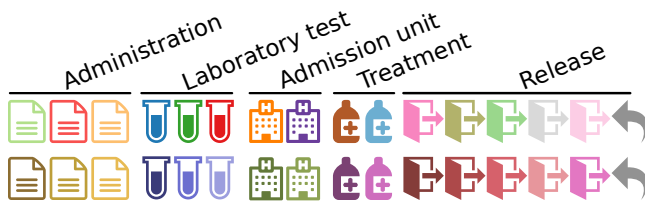
Figure 4: Icons and color palettes that can be found in SepVis. The first palette is intended to facilitate users to spot individual activities. On the other hand, the second one aims to enable users to identify groups of activities.

Color palette. In order to give a color to each activity, we offer two different approaches in our tool. The color palette can be switched easily and the mechanism to do so is explained in Sect. 6.1. We consider that colors can play an important role in two different situations. On the one hand, they can help to identify groups of activities without focusing on the activity itself. On the other, they can be used to spot specific activities more easily. In the first case, the color palette is divided by categories of activities, giving a color to each category and a specific tone of that color to each activity within the category. In the second case, we use a qualitative scheme of colors. In this way we allow users to quickly find activities of interest since they all have different colors. Both color palettes can be seen in Fig. 4.

Shifting & opaqueness. Due to the fact that several events can happen at the same point in time causing overlap of icons, we shift and make icons opaque. Shifting helps to avoid two icons from being placed at exactly the same position. In our tool, we apply a small constant shifting to all icons. In this way, we ensure that no full overlap occurs between two consecutive icons. On the other hand, opaqueness ensures that the icon placed on top of another one is always visible. This helps to disambiguate overlapping icons resulting in a better appreciation of them and their distribution over time.

Border. In some cases, the border of the icons is used to present a special situation such as a model move. Recall that a model move is a situation in which something should have happened in reality, but it did not. This is important since it can reflect a missing step in a process, e.g., the administration of a medicine or the request of a laboratory test. We use this codification since it does not clutter the view excessively and the model moves can still be seen.

5.2 Activity links design

The activities within a trace are considered to occur sequentially in time. In order to reflect this in an explicit way, our tool uses links between activities. Besides this, those links can be used to encode some extra information. In our approach, links can represent three different states (see Fig. 5):

Synchronous moves. This is the normal and most common case. It reflects the link between two activities that happened both in reality and according to the process model. This type of link is presented in our tool by a straight line with a grey color connecting two activities.

Log moves. These moves reflect an abnormal behavior that was recorded in the event log and does not match the process model. We use a red, thicker line in order to display this type of move. Since the line links two activities, it is colored depending on the type of the latter activity. If that activity is marked as a log move, then the line will be colored as red.

Figure 5: Three types of lines encoding three different types of moves: synchronous moves, log moves and model moves, respectively from top to bottom.

Model moves. The last type of move depicts a case in which something should have happened according to the process model, but it did not. A dotted line is used in this case. The reason behind this choice is that we try to highlight the fact that we do not know if this activity happened in reality and it was not recorded, or that it did not actually occur.

6 SEPVIS

In this section we describe SepVis by explaining its components and how they interact. Our approach consists of clinical pathways exploration, filtering and attribute selection. An overview of our approach can be seen in Fig. 6. Each view interacts with the others in order to provide an intuitive perception of patients' pathways to support the tasks **T1-T3**.

From the task analysis we derived the core components of SepVis. An overview of patients pathways as well as the most and the least frequent patterns **(T1)** over time are given in our main view, the *Pathways view*. For a life-threatening condition such as sepsis, diagnosis time and treatment occurrence play a crucial role in patients' survival. Hence, time and filtering options by activity/pathways duration need to be provided for a proper comprehension of patient management and resource efficiency **(T2)**. Time information is encoded in the *Pathways view* and data can be filtered from the *Performance search*. Clinical data **(T3)**, such as recorded conditions when the patients enter in the hospital, the age and the exams they underwent before entering in a normal care unit, is integrated with the application. The user can use the *distributions panel* to create a cohort of patients and immediately recognize patterns in the *Pathways view*. In the following subsection, we illustrate these visualizations and their interaction in more detail.

6.1 Pathways View

This is the core view of SepVis and it consists of icons representing activities of the entire event log. Those activities are linked by lines in order to explicitly encode the temporal relation between two consecutive activities. The sequence of icons and lines within a row shapes a trace. Contrary to the current process centric visualizations (see Fig. 3), we distribute activities over time. The icons that belong to a particular trace are placed along the horizontal axis according to their accumulated duration since the starting activity of the trace, which is always considered to have an accumulated length of zero. Because we cope with data regarding a life-threatening condition where actions need to be taken in a few hours from the entrance of the patient in the hospital, the horizontal axis of this view represents time intervals with different scales. Following the guidelines [14], we opt for the following intervals: zero to 1 hour, 1 to 3 hours, 3 to 12 hours, 12 to 24 and then to 48 hours, 1 to 2 weeks and then onwards. Thus, this view comprises seven scales along the axis (one for each interval). All intervals are given a different horizontal space. The first intervals are wider to enable to focus on the events occurring in the first moments of patients hospitalization and the first taken actions.

The lines in this view also encode deviation information, which is provided by the process mining alignment algorithm. Typically, we can find two different types of deviations: log and model deviations (see also Sect. 2 and 5). The first deviation is encoded by coloring the line between two activities with a red color and increasing
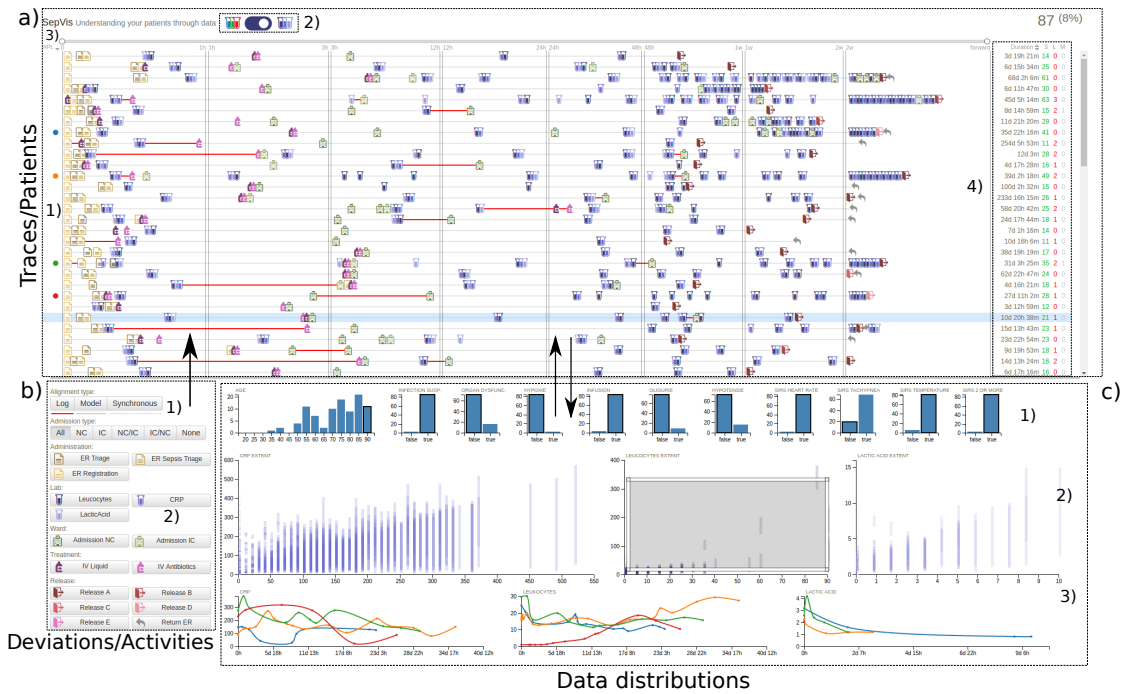
Figure 6: The components and interactions within SepVis: the *Pathways view* (a) shows all the pathways of patients, which are grouped by variants over time, the activity/deviations filtering panel b) and the clinical attributes distribution panel c). Details about performances are shown in a.4). Distributions of the clinical attributes for the current selection of patients are shown in c.1) by the orange bins. Distributions of the extents of the laboratory tests are shown in c.2). Single-patients trends, such as treatment values, are displayed in c.3).

Figure 7: Example of an expanded variant which originally contains 11 traces in total and 4 of them match the current filter setting. A selected trace is indicated by the blue dot at the left of the trace.

the thickness of the line. Because a line links two activities, we color lines as red when the last activity occurs as a log move. The second deviation is depicted by using a dotted line(s) between the activity that makes a model move and the surrounding activities. This approach is susceptible to be cluttered because of icon overlaps. We solve this problem by shifting the icons when necessary, that is, when the link between two activities encodes meaningful information (i.e., deviations).

The traces are grouped in variants, which represent the same sequence of activities for all the traces that they group. This grouping just happens when there is more than one trace of a given variant. The visualization of the variants is slightly different from the visualization of the individual traces. It collapses all the traces which belong to that variant and just shows the variant itself together with two numbers: one indicating the total number of traces for the variant, and another one showing the number of traces that match the current filtering setting. Moreover, the placement of the icons within a variant is based on the average accumulated duration, which is computed over all the traces that form a variant.

The variants can be sorted by different criteria by clicking on one of the fields on the header of the view. These criteria consist of a number of traces per variant, total (average) duration of the traces

(variants) and delay between two selected activities.

The right side of this view displays a summary of how many deviations and normal moves a specific variant or trace contains. These figures are shown by using different colors as follows: synchronous moves are given in green, log moves in red and model moves in grey. This piece of information gives a quick summary of how well the trace fits the process model.

For each row in the visualization, four main interactions are available:

Clicking a variant. This interaction displays all the traces that belong to that group with a drop-down effect. The user can then explore each individual case. Fig. 7 shows an example of an expanded variant. The traces that belong to the variant are placed beneath it and the variant is highlighted with a bluish color. A vertical thick line at the left of all the traces is used to highlight that they belong to the same variant and differentiate from the rest. When clicking on a variant, the data attributes of all the traces that it contains are reflected in the charts. This feature is explained in more detail in Sect. 6.4.

Hovering activities. Mousing over either icons or lines opens a tooltip, which shows information about the hovered element. When the hovered element is a line, it simply shows the elapsed time between the two linked activities. When it is an icon, it shows some basic information such as activity name, time elapsed since the beginning of the trace and time elapsed since the previous activity. Moreover, when the icon represents a more meaningful activity such as a lab test, the tooltip also includes information about the value of the hovered activity, or the minimum and maximum values for that test when hovering a variant. Finally, in case the hovered activity represents a log or model move, the tooltip also shows this information.

Hovering traces. Users can also mouse over traces that represent

particular care flows, in which case the views depicted in Sect. 6.4 are updated showing the interesting values for the specific case. A more complete description of this interaction is given later.

Selecting activities. We enable the user to select singular activities. This interaction is described in Sect. 6.2.

Furthermore, the user can directly manipulate the view by the top slider (Fig. 6(a.3)). The behavior of this slider is explained in Sect. 6.3.

On the header of this view we can find different components. The upper right corner of the screen displays the number of patients that match the current filter setting besides the percentage of the number of patients. This information gives an idea of how representative the current population is. Also, on top of the view (Fig. 6(a.2)) we can see the widget used to switch between the different color schemes that were depicted in Sect. 5.1. This component is presented as a switch. At both sides of the switch we can find an example of icons that are colored with different color schemes. This indicates which color scheme we selected. Once the user selects one color scheme, the view updates to present the icons with the new color scheme.

6.2 Activity filters and conformance checking

In this section we describe the behavior of the tool regarding filtering traces by activity or sequence of activities, and conformance information, that is, deviations in the process. This component can be seen in Fig. 6(b).

Traces can be filtered out by the types of deviations they contain. Our tool provides four different options: all cases, only log moves, only model moves and only synchronous moves. This is useful to retrieve traces that entirely fit the process model, or those ones that have specific deviations. This directly relates to task **T3**.

Also, traces can be filtered by activities by interacting with the buttons (see Fig. 6(b.2)). This feature is provided to accomplish task **T2**. Clicking on a button in the activity box filters out the traces that do not contain such activity. Once this happens, the main view automatically reacts by decreasing the opacity of the other activities in order to highlight the selected activity. Furthermore, the information displayed at the right side of the main view automatically updates and shows the time elapsed until that activity occurs for each trace. In case of variants, it shows the average elapsed time. Users can then sort traces by this new value.

This feature is extended by enabling users to select a second activity. In this case, the main view filters out the traces that do not contain both activities in the specified order, highlights both selected activities, creates lines between the two selected activities and eliminates the other ones to present a view more focused on the selected activities. Also, similar to the case of selecting one activity, the main view shows the time elapsed between the first occurrence of the first selected activity and the first occurrence of the second selected activity. Traces can then be sorted by this new value.

The new information displayed in both situations can be used by the other components of the application to filter out traces and is explained in Sect. 6.3. This component also plays the role of legend in our application since it shows all the activities with their colors and their names.

An extra filtering mechanism is provided with regard to activity filtering. This filter enables users to explore care pathways in which patients were admitted to NC, IC, NC and then IC, and IC and then NC units. In addition, the option to explore patients which went to neither NC nor IC is provided. These options address task **T1**.

6.3 Temporal filters: Performances

The slider located above the main view is aimed to filter out traces according to some time criterion. The criterion depends on the current context of the tool. Initially, this component just considers
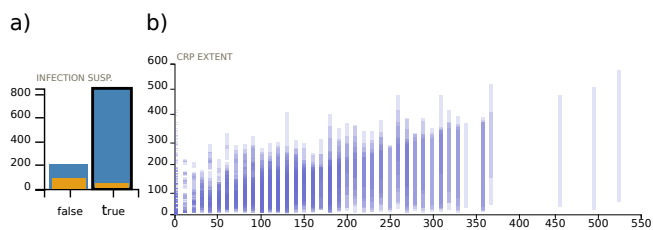

Figure 8: a) Bar chart showing three different pieces of data: the values for all the traces that match the current filters displayed as blue bars, the values for all the traces of the selected variants displayed as golden bars and the value for *infection suspected* attribute of the hovered trace displayed as a black border. b) Chart that aggregates patients with similar values for the CRP tests.

the total duration of the traces. Thus, users can shift both ends of the slider in order to set minimum and maximum thresholds for the duration of the traces in which they are interested. The slider follows the same time scales as the main view.

If an activity is selected as described in Sect. 6.2, the lower and higher values of the slider apply only to the first and last occurrence of the selected activities. Therefore, if users select only one activity, the slider will consider those traces in which the first occurrence of such activity happens after the lower value. If users select two activities, then both lower and higher thresholds will be considered to filter out those traces in which the first occurrence of the first activity happens after the lower value and before the higher value of the slider.

6.4 Attribute filters: Cohort distributions

The attributes panel is provided to look in more detail at patient's pathways from the clinical perspective. Among the available attributes we can find the age of the patients and other clinical attributes of their conditions at the ER registration. A more complete description of the dataset is given in Sect. 4.

Clinical attributes are shown as individual, interactive bar charts. Interacting with one of the bar charts triggers an update from the rest of charts and views of the tool. Also, these charts are linked with the main view when a variant is selected and a trace is hovered. When this happens, two outcomes are possible (see Fig. 8(a)):

- If a variant was selected, the data from all the traces that belong to it is displayed by showing overlapping bars for each bar-chart. These bars are horizontally shrunken and positioned on top of the corresponding bars. Moreover, the color of these bars is different to easily spot the difference between the whole data and the selected traces.

- If a trace was hovered, then the bars with the values of the data of that specific trace are highlighted by adding a black stroke. This way we can explore the attributes for particular traces.

The data view also includes charts to present the values for different laboratory tests that are available in the dataset. Due to the fact that there is more than one value per test, we opt for using two charts per type of test: one chart that shows a summary of the values for all the traces currently in the main view and another one which displays all the values of that test for traces selected by the user.

The first type of chart (see Fig. 8(b)) provides a representative picture of the range of values for a specific laboratory test. In our approach, we use a histogram in which we group patients by the difference between the global maximum and minimum values for the laboratory test. The horizontal axis is divided into slots of the same size (10 units in the example), and each data item (patient) is horizontally assigned to the closest slot based on the difference of the extreme values. Each patient is depicted by a bar. The height

of a bar encodes the variation of the value for the laboratory test. Moreover, each bar is assigned a semitransparent filling color. This representation enables to detect where the trend of the distribution of values is and the existence of outliers at the same time, while keeping the whole picture of the data. Specific parts of the data can be selected by brushing over the chart, making the rest of the tool react to show just the selected patients.

In the bottom part of SepVis (see Fig. 6), we can find the charts which display all the values of a particular test for the selected traces. This type of chart is aimed to give a better understanding of the evolution of the patient over time. Therefore, the horizontal axis of this type of chart encodes time and the vertical axis encodes the values of the test. This type of chart only offers the interaction to see the value of specific points by mouse hovering them.

These charts, and the attributes that they depict, are interesting for domain experts that are aware of guidelines since normal/abnormal values might indicate some decisions that may need to be taken. With the interactions described in this section, we enable users to explore the distribution of these attributes in a proper way.

6.5 Implementation

SepVis is implemented as a web application and it is entirely programmed in D3.js v4 [3]. We use Bootstrap v3 for a base layout, in combination with Icon Moon icons. Finally, we make use of several JavaScript libraries: jQuery v1.12, underscore v1.8 and Crossfilter.js v1.3. Finally we use color palettes from D3.js that were taken from ColorBrewer.

Regarding process mining technologies, we use the alignment technique described in [21] which is available as a plugin in ProM [22].

7 USE CASES AND RESULTS

We demonstrate the usefulness of SepVis by some use cases with questions extracted from conversations with doctors involved in the project [10] and the guidelines provided by the Surviving Sepsis Campaign [9] released in 2012.

7.1 Clinical pathways in NC and IC units

In an emergency department for suspected sepsis patients, one of the first questions that the clinical personnel tries to answer concerns clinical pathways across the treatment units.

For instance, it is interesting to know the flow of patients that are only admitted to the NC ward, only to the IC unit, or first to the normal care unit and, only then, to the intensive care unit. The last category is considered of great interest by clinical personnel because it hints to patients whose conditions have worsened after being admitted. In these situations, fast decisions need to be taken. Given the high interest in these particular activities, we provide a dedicated button bar for selecting these particular groups: patients that entered respectively only in the Normal Care unit, Intensive Care unit, Normal Care and then after Intensive Care unit, or the other way around. A special group is also included for patients who did not enter neither Normal Care nor Intensive Care. This group of patients might reflect those who went to the emergency room and did not need to be monitored.

Our dataset has 67% of patients that went only to the NC unit; 23% of them did not enter in *any* of the two units during their stay at the ER. Selecting the latter option, the users can notice that, despite the fact that these should be the cases with the shortest duration, some patients activities have been recorded even after one day. In total, the number of patients that present deviations (model moves and log moves) were around 48% of the total cases. Of this group, eleven patients are likely matching the diagnostic criteria for a sepsis condition. The diagnosis of an inflammatory state is typically performed by assessing the manifestations of the Systemic Inflammatory Response Syndrome (SIRS), such as alterations of body temperature or increase of heart rate and white blood cells. Therefore, we can look for patients with a clear need of specific care by filtering with SIRS-2 criteria (patients matching two or more SIRS criteria), suspected infection and positive organ dysfunction, the major indicator of a sepsis condition [6]. The user gets a list of patients (1% of the total patients). For some reason, perhaps mistakes in recording patients activities or other decisions, these patients have not been treated in the specific care units. Switching back to *all* patients, we get around the 7% the cases whose treatments is clearly lasting more days because of follow-up and monitoring tests.

7.2 Performances and delays

Guidelines for life-threatening conditions provide clear recommendations regarding treatment administration time. Early identification of septic patients is crucial to prevent delays in management and treatment. Many studies focus on the impact of time to antibiotics administration in respect to triage screening tools and protocols [8, 15, 19]. Current guidelines in management of sepsis patients [6, 9] state that the distribution of antibiotics to patients in sepsis conditions needs to take place within the first hour from a sepsis triage. Due to lack of resources or delays this might not happen, leading to serious risk to the patients, which has to be avoided. In SepVis, the users can get insight into these circumstances. Firstly, they select the 'suspected' sepsis patients according to the clinical attributes at disposal. Next, the user chooses the *ER-Sepsis Triage* event as first option and *antibiotics* as second from the activity list. Thus, we get the resulting patients in the *pathways view* and the elapsed time between two activities. The user can observe that many patients actually received antibiotics between 1 and 3 hours and after 3 hours. Most likely these patients were not considered to be in a critical condition, therefore the administration was delayed or simply recorded after some time. In this view we can also see some deviations in the clinical pathway. This is because of the proposed protocol used in SepVis, which states that *intravenous antibiotics* is a treatment which is followed or proceeded by also *intravenous liquids* administration. In some cases, it seems that this administration of liquids was not performed or recorded.

Another interesting case is visible by considering the *3-hours bundle* of [9]. In order to simplify and improve the care of patients with severe sepsis, the *Surviving Sepsis Campaign* released bundles, which are a selected set of elements to be considered and executed to aim at the best outcome. In the 3-hours bundle, patients with hypo-tension or lactate values below 4 mmol/L need to be treated with *intravenous fluids* in order to expand their circulating volume. The user can check the patients matching the hypo-tension attribute and lactate values above the threshold. In our scenario, the dataset comprises four patients that got IV fluids after 3 hours. In combination with EHRs, this kind of result can be explored in more depth. Unfortunately, the accuracy of the clinical attributes at our disposal it is not sufficient to draw conclusions about the outcome of these suspected delayed interventions.

7.3 Deviations and clinical attributes

As we described in the Background section, in process events, two different kinds of deviations can be detected by conformance checking: log-moves and model-moves.

With SepVis, we enable the users to see when and where these deviations occur. From the clinical perspective, a log move can be described as a decision that does not match the expected behavior. A model-move instead, might indicate forgotten actions, missed decisions, lack of resources and such.

Thus, we enable the user to filter out patients that matched entirely the proposed model. At this point, deviating cases can be analyzed over time and by clinical attribute distributions. Once the deviating cases are available, the user might be interested to see single patients'

trends in test values over their stay in the hospital. The user can select a few patients from the deviating case list. Then a click on a test activity such as *CRP*, *leukocytes* or *lactic-acid* levels, triggers the update of the respective charts showing the fluctuations over time.

8 CONCLUSION AND FUTURE WORK

In this work we have proposed a visualization approach that enables the analysis of process centric information of an event-log in combination with clinical longitudinal data.

We have shown the usage of the current implementation in accordance to standard clinical tasks and relevant insights in respect to current sepsis management guidelines. Our design comprises a single interface, which is compact and more intuitive in comparison to current solutions in the process-mining domain. It provides the necessary components to interact with cases and at the same time analyze their clinical pathways filtering by time, performances and clinical data.

The core components of SepVis allow the exploration of an event log by interacting with its main elements: cases, activities, and attributes over time. While we demonstrated SepVis on an example dataset for sepsis, our approach is generic and can be used with other event logs. The current limitation is given by the fact that our visual interface has been built for an event-log with a low number of activities (16) whereas processes may comprises hundreds of different activities. At present, SepVis exclusively supports a single-cohort investigation per time.

In future work we aim to enable the user to work in terms of cohorts of patients and present correlations, distributions and process deviations between them. We also aim to increase the scalability of the design in order to deal with larger and more complex datasets.

REFERENCES

[1] R. A. Balk. Systemic inflammatory response syndrome (SIRS): where did it come from and is it still relevant today? *Virulence*, 5(1):20–6, jan 2014. doi: 10.4161/viru.27135

[2] D. Borland, V. L. West, and W. E. Hammond. Multivariate Visualization of Longitudinal Clinical Data. *Proceedings 2016 IEEE VIS Workshop on Visual Analytics in Healthcare, Chicago,IL*, 2016.

[3] M. Bostock, O. Vadim, and H. Jeffrey. Data-Driven Documents. *IEEE Transactions on Visualization and Computer Graphics*, 17:2301 – 2309, Dec 2011.

[4] Celonis. [Online; accessed 15-May-2017].

[5] A. S. C.Plaisant, R. Mushlin. LifeLines: Using Visualization to Enhance Navigation and Analysis of Patient Records. *American Medical Informatic Association Annual Fall Symposium*, pp. 9–11, 1998.

[6] R. P. Dellinger, M. M. Levy, A. Rhodes, D. Annane, H. Gerlach, S. M. Opal, J. E. Sevransky, C. L. Sprung, I. S. Douglas, R. Jaeschke, T. M. Osborn, M. E. Nunnally, S. R. Townsend, K. Reinhart, R. M. Kleinpell, D. C. Angus, C. S. Deutschman, F. R. Machado, G. D. Rubenfeld, S. Webb, R. J. Beale, J.-L. Vincent, R. Moreno, and Surviving Sepsis Campaign Guidelines Committee including The Pediatric Subgroup. Surviving Sepsis Campaign: International Guidelines for Management of Severe Sepsis and Septic Shock, 2012. *Intensive Care Medicine*, 39(2):165–228, feb 2013. doi: 10.1007/s00134-012-2769-8

[7] P. M. Dixit, H. S. G. Caballero, A. Corv, B. F. A. Hompes, J. C. A. M. Buijs, and W. M. P. van der Aalst. Enabling interactive process analysis with process mining and visual analytics. In *Proceedings of the 10th International Joint Conference on Biomedical Engineering Systems and Technologies - Volume 5: ACP, (BIOSTEC 2017)*, pp. 573–584. INSTICC, ScitePress, 2017. doi: 10.5220/0006272605730584

[8] M. Francis, T. Rich, T. Williamson, and D. Peterson. Effect of an emergency department sepsis protocol on time to antibiotics in severe sepsis. *CJEM*, 12(4):303–10, jul 2010.

[9] M. D. Howell, A. M. Davis, and S. G. Management of Sepsis and Septic Shock. *JAMA*, 317(8):847, feb 2017. doi: 10.1001/jama.2017.0131

[10] F. Mannhardt. Sepsis Cases - Event Log, jan 2016. doi: 10.4121/UUID:915D2BFB-7E84-49AD-A286-DC35F063A460

[11] F. Mannhardt and D. Blinde. Analyzing the trajectories of patients with sepsis using process mining. In *RADAR+EMISA 2017*, vol. 1859 of *CEUR Workshop Proceedings*, pp. 72–80. CEUR-WS.org, 2017.

[12] M. Monroe, R. Rongjian Lan, H. Hanseung Lee, C. Plaisant, and B. Shneiderman. Temporal Event Sequence Simplification. *IEEE Transactions on Visualization and Computer Graphics*, 19(12):2227–2236, dec 2013. doi: 10.1109/TVCG.2013.200

[13] M. Monroe, K. Wongsuphasawat, C. Plaisant, B. Shneiderman, J. Millstein, and S. Gold. Exploring point and interval event patterns: Display methods and interactive visual query, 2012.

[14] Guidelines for the management of sepsis (including neutropenic sepsis). [Online; accessed 15-May-2017].

[15] C. Patocka, J. Turner, X. Xue, and E. Segal. Evaluation of an Emergency Department Triage Screening Tool for Suspected Severe Sepsis and Septic Shock. *Journal For Healthcare Quality*, 36(1):52–61, jan 2014. doi: 10.1111/jhq.12055

[16] D. S. Pieczkiewicz, S. M. Finkelstein, and M. I. Hertz. Design and evaluation of a web-based interactive visualization system for lung transplant home monitoring data. In *AMIA Annual Symposium Proceedings*, vol. 2007, pp. 598–602. American Medical Informatics Association, oct 2007.

[17] A. Shih and S. C. Schoenbaum. Measuring Hospital Performance: The Importance of Process Measures. *The Commonwealth Fund*, July 2007.

[18] J. V. Sutherland, W.-J. van den Heuvel, T. Ganous, M. M. Burton, and A. Kumar. Towards an intelligent hospital environment: OR of the future. *Studies in health technology and informatics*, 118:278–312, 2005.

[19] S. Trzeciak, R. P. Dellinger, N. L. Abate, R. M. Cowan, M. Stauss, J. H. Kilgannon, S. Zanotti, and J. E. Parrillo. Translating Research to Clinical Practice. *Chest*, 129(2):225–232, feb 2006. doi: 10.1378/chest.129.2.225

[20] W. van der Aalst. Data Science in Action. In *Process Mining*, pp. 3–23. Springer Berlin Heidelberg, Berlin, Heidelberg, 2016. doi: 10.1007/978-3-662-49851-4_1

[21] W. van der Aalst, A. Adriansyah, and B. van Dongen. Replaying history on process models for conformance checking and performance analysis. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 2(2):182–192, mar 2012. doi: 10.1002/widm.1045

[22] B. F. van Dongen, A. K. A. de Medeiros, H. M. W. Verbeek, A. J. M. M. Weijters, and W. M. P. van der Aalst. The ProM Framework: A New Era in Process Mining Tool Support. In *Applications and Theory of Petri Nets 2005*, pp. 444–454. Springer, Berlin, Heidelberg, 2005. doi: 10.1007/11494744_25

[23] W. Widanagamaachchi, Y. Livnat, P.-T. Bremer, and V. Pascucci. Interactive Visualization and Exploration of Patient Progression in a Hospital Setting. *Proceedings 2016 IEEE VIS Workshop on Visual Analytics in Healthcare, Chicago, IL*, 2016.

[24] K. Wongsuphasawat and D. H. Gotz. Outflow: Visualizing patient flow by symptoms and outcome. *Proceedings IEEE VIS 2011 Workshop on Visual Analytics in Healthcare*, 2011.

[25] K. Wongsuphasawat, J. A. Guerra Gómez, C. Plaisant, T. D. Wang, B. Shneiderman, and M. Taieb-Maimon. LifeFlow: Visualizing an Overview of Event Sequences. *Proceeding CHI '11 Proceedings of the SIGCHI Conference on Human Factors in Computing Systems Pages 1747-1756*, 2010.