

# Process Mining the Trauma Resuscitation Patient Cohorts

Sen Yang, Fei Tao, Jingyuan Li, Dawei Wang, Shuhong  
Chen, Ivan Marsic  
Rutgers University  
Piscataway, NJ  
{sy358, ft110, jl2056, dw514, sc1624, marsic}@rutgers.edu

Omar Z. Ahmed, Randall S. Burd  
Children's Nat'l Medical Center  
Washington, DC  
{oahmed3, rburd}@rutgers.edu

**Abstract**—In this study, we present a framework for analyzing associations between patient cohorts and the trauma resuscitation procedures their patients received. Our framework works by quantifying associations between discovered patient cohorts and treatment patterns. We evaluated our framework on a trauma resuscitation dataset collected in a level 1 trauma center. Our experimental results show that using weights learned by our algorithm improves measurements of patient similarity. Four patient cohorts were then found via clustering, and statistically significant resuscitation patterns were discovered using process mining techniques. Though only tested on the trauma resuscitation process, our framework can be generalized to analyze other medical processes.

**Keywords**—Process Mining; Patient Cohorts Analysis; Trauma Resuscitation; Medical Workflow Analysis.

## I. INTRODUCTION

Patient cohort analysis is widely used to make clinical discoveries in medical research [1]-[4]. A patient cohort is defined as a group of patients who share similar context attributes. In trauma resuscitations for example, trauma patients of a same cohort share similar attributes such as demographics (e.g., age, gender, ethnicity, insurance, and medical history), injury information (e.g., injury type, injury severity, and injury area), and trauma attributes (e.g., day vs. night shift, trauma activation level and pre-arrival notification). In traditional patient cohort analysis [2][3][4], medical analysts define patient cohorts according to targeted attributes selected by medical experts; other context attributes were considered confounding and ignored. Oriented by domain knowledge, these studies were likely to reveal expected results but miss findings unfamiliar to experts.

Process mining [1] is another set of analytical techniques that has been recently applied to medical process analysis. It has been used to discover medical process models [6], measure compliance of process executions to expert models [7], and analyze medical process deviations [8]. Existing process mining research [5] however, mostly mines knowledge from entire process datasets, without studying the differences among subsets of the data.

In this paper, we present a framework for analyzing medical process data by combining both process mining and patient cohort analysis. Our medical process data consists of process activity logs (e.g., trauma resuscitation executions) and context attributes (e.g., patient demographics) associated with each process case. Our framework works in three steps.

First, it applies data exploration methods on patient attributes to find data-driven cohorts. Second, it discovers process patterns (e.g., treatment patterns) from activity logs using process mining techniques. Third, it tests the significance of the correlations between process patterns and patient cohorts.

We applied our framework on a real-world medical process: the trauma resuscitation. The trauma resuscitation is a fast-paced process, where multidisciplinary teams need to rapidly identify and treat potentially life-threatening injuries in an injured patient. Analyzing correlations between treatment executions and patient cohorts can potentially improve our understanding of the process and hopefully improve patient outcomes.

Our contributions in this study are:

- A framework for discovering and analyzing the associations between trauma patient cohorts and trauma resuscitation procedures. Our framework is easy to implement and can be used for analyzing processes with event (or activity) logs and external context attributes.
- A practical algorithm and experimental procedure to learn the weighted importance of attributes with little human intervention. Unit weights are usually assigned to attributes when calculating data similarity for clustering, as the actual significance of each attribute is unknown. In this study, we designed an experiment to very efficiently acquire medical experts' input to supplement attribute weight learning.
- An analysis of statistically significant correlations between context attributes (aggregated as patient cohorts) and discovered medical treatment patterns in a real-world dataset of 123 trauma patients.

## II. DATA DESCRIPTION AND FORMALIZATION

One hundred and twenty-three trauma resuscitation videos were collected from the trauma bay of Children's National Medical Center, Washington DC. The videos were reviewed jointly by a surgeon with Advanced Trauma Life Support (ATLS) [9] certification and trauma clinical nurse specialists to encode the activity traces (Table I (A)). A total of 7154 main activities of 44 types were selected for this study. Twenty-six context attributes were collected from the trauma database or from medical chart review, including patient age, gender, trauma activation level, mechanism of injury (penetrating, blunt, burn, etc.), date and time of patient arrival, Injury Severity Score (ISS), Glasgow Coma Score (GCS),

intubation status, and Abbreviated Injury Scale (AIS) (Table I (B)). The collection and use of the data for this study were approved by the hospital’s Institutional Review Board.

Here we first define terms and notations. The process  $\mathbf{log} \mathbf{L} = [c^{(1)}, \dots, c^{(l)}]^T$  is a vector of elements  $c^{(i)}$ . Each  $c^{(i)} = \{id^{(i)}, \mathbf{x}^{(i)}, \mathbf{T}^{(i)}\}$  (Table I (D)) representing a unique **case**, which is indexed with a unique case  $id$ , contains the activity trace  $\mathbf{T}^{(i)}$ , and has a vector  $\mathbf{x}^{(i)}$  of context attributes. An activity **trace** is  $\mathbf{T}^{(i)} = [a_1^{(i)}, \dots, a_m^{(i)}]^T$ , where  $m$  total activities  $a$  are ordered by activity start time. Traces of different executions may have varying lengths because complex processes may contain optional, omitted, or even erroneously performed activities. **Context attributes**  $\mathbf{x}^{(i)} = [x_1^{(i)}, \dots, x_n^{(i)}]^T$  is a vector of  $n$  recorded patient demographics, injury information, and trauma attributes.

### III. METHOD

In this section, we describe the core techniques we used in attribute weight learning, patient cohort discovery, process mining, and statistical analysis. We learned the attribute weights to decide the importance of different context attributes. With this information, we can find more accurate patient cohorts through clustering algorithms. We then mined the treatment patterns within each patient cohort and analyzed them with statistical methods.

#### A. Attribute Weight Learning

In our framework, patient cohort is decided by unsupervised clustering. The clustering performance is highly associated with the attributes used. The discovered patient cohorts may be meaningless if irrelevant or unimportant attributes are used. Due to lack of any prior knowledge, the attribute weights are typically set to unit weights (i.e., all attributes have the same weight as one). With domain knowledge available, it is possible to obtain the attribute weights by asking medical experts to provide a score (e.g., in scale of 0 – 10) for each attribute. This approach however is difficult to apply in practice. We tried this method by asking the medical experts in our team to decide a dictionary of attribute weights. Our medical experts later deemed it too arbitrary to manually set quantified weights across all attributes. Even if the medical experts could provide a set of weights, it would likely heavily bias the results towards their domain knowledge, leading to an “expected” result. Hence we designed a simple experiment to collect medical decisions and developed a learning algorithm for learning attribute weights.

Our experiment used 41 sets (denoted as  $\mathbf{S}$ ) of three patients each (e.g., Patients A, B, C in Table II), randomly drawn (without replacement) from the trauma resuscitation dataset. A surgeon was asked to decide the most similar pair among the three patients, (A, B), (A, C) or (B, C), based on their context attributes only. They used their domain knowledge to judge how important the differences of attributes were, and decided which pair of patients were more likely to be in a same cohort. In our example (Patients A, B, C in Table II), our medical expert labelled patient pair (B, C) as the one most likely to be observed in a same cohort. This experiment is simple and does not require much human effort.

TABLE I. ACTIVITY TRACE (A), CONTEXT ATTRIBUTES (B), DATA STATISTICS (C) AND DATA FORMALIZATION (D).

Case ID	Activity	Start Time	End Time	Case ID	xxx1	xxx2
xx1	Pt arrival	0:00:00	0:00:01	Age category	24-96	24-96
xx1	Visual assessment-AA	0:00:45	0:00:52	Sex	Male	Female
xx1	Chest Auscultation-BA	0:00:55	0:00:58	Night Shift	0	1
xx1	R DP/PT-PC	0:01:04	0:01:05	Weekend	0	0
xx1	Total Verbalized-GCS	0:01:29	0:01:30	Pre-arrival Notification	1	0
xx1	Total Verbalized-GCS	0:01:50	0:01:51	Trauma Activation Level	Transfer	Attending
xx1	Right pupil-PU	0:02:12	0:02:18	Intubation	0	0
xx1	Left pupil-PU	0:02:19	0:02:24	Glasgow Coma Score >13	1	0
xx1	Right pupil-PU	0:02:24	0:02:25	Injury Type	Blunt	Penetrating
xx1	Visual inspection-H	0:02:33	0:02:34	Injury Severity Score	5	12
xx1	Palpation-H	0:02:33	0:02:37	Neck Injury Severity Score	3	5

(A) Trauma resuscitation trace		(B) Context attributes		
Properties	Stats	ID	Ext. Attributes	Resus. Traces
Num. Cases (or Patients)	123	$id^{(1)}$	$x_1^{(1)}, \dots, x_g^{(1)}$	$a_1^{(1)}, \dots, a_k^{(1)}$
Num. Total Activities	7154	$id^{(2)}$	$x_1^{(2)}, \dots, x_g^{(2)}$	$a_1^{(2)}, \dots, a_k^{(2)}$
Num. Activity Types	44	$\vdots$	$\vdots$	$\vdots$
Num. External Attributes	18	$id^{(n)}$	$x_1^{(n)}, \dots, x_g^{(n)}$	$a_1^{(n)}, \dots, a_k^{(n)}$
Data Collection Time Period	2014.08 – 2016.12			
Size of Medical Team	[7, 12]			
Longest Trace (Num. Acts.)	110			
Shortest Trace (Num. Acts.)	26			
Avg. Num. Acts. in Traces	58.6			

(C) Data statistics	(D) Data formalization

We then used these labeled results (denoted as  $\mathbf{P}_{Dr}$ ) as inputs to our attribute learning algorithm (Alg. 1).

Our learning algorithm was designed with the core idea that by adjusting the weights of context attributes, we can increase the classification accuracy (i.e., deciding which pair of patient is more similar). The similarity measure is defined using weighted Euclidean distance [10][11]:

$$d_{AB} = \left( \sum_{i=1}^n w_i (x_i^{(A)} - x_i^{(B)})^2 \right)^{\frac{1}{2}} \quad (1)$$

where  $w_i$  is the weight given to the  $i$ -th component.  $x_i^{(A)}$  and  $x_i^{(B)}$  are the  $i$ -th context attributes of patients A and B. The categorical variables were converted into numeric by creating dummy variables. If distance  $d_{AB} < d_{AC}$  &&  $d_{AB} < d_{BC}$ , it means that patient A and B were the most similar pair given a set of weights  $\mathbf{w} = [w_1, \dots, w_n]^T$ . Then if the label given by medical expert is also pair (A, B), it is a hit, otherwise a miss. The overall classification accuracy over  $N$  patient sets is defined as the ratio of hits:

$$acc = \frac{num(hits)}{N} \quad (2)$$

Our algorithm updates the attribute weights iteratively. At each iteration, we test adding (Step 4) or subtracting (Step 10) a unit weight from a single attribute weight  $w_i$ . An important boundary condition is  $w_i \geq 0$  (Step 11), otherwise  $w_i$  does not have physical meaning in similarity calculation. Then we calculate the updated accuracy (Step 6 & 13) after addition and subtraction. Last, we update (Step 17-23) the attributes which lead to the highest accuracy (Step 16). The algorithm terminates when the accuracy stays unchanged for a defined number of iterations (Step 24). Alg.1 is based on greedy search [12]. At each step, we only update the weights on attributes which provide maximum improvement. Our

**Algorithm 1. Patient Attribute Weighting**

**Input:**  $N$  random drawn sets  $\mathcal{S}$ ; labels from medical expert  $P_{Dr}$   
**Output:** Learnt Attribute weights  $\mathbf{w}$

Step 1. Initialize  $\mathbf{w} \in \mathbb{R}^{1 \times n}$ , and  $\mathbf{acc} \in \mathbb{R}^{1 \times 2n}$  as vectors of zeros  
Step 2. **do**  
Step 3.     **for** each weight  $w_i \in \mathbf{w}$ , **do**  
Step 4.      $w_i++$   
Step 5.     Calculate the most similar pair  $P_s$  in each set of  $\mathcal{S}$   
Step 6.     Calculate  $acc_i$  ( $acc_i \in \mathbf{acc}$ ) based on  $P_{Dr}$  and  $P_s$   
Step 7.      $w_i--$   
Step 8.     **end for**  
Step 9.     **for** each  $w_i \in \mathbf{w}$ , **do**  
Step 10.      $w_i--$   
Step 11.     **if**  $w_i < 0$ ,  $w_i++$ , **continue**  
Step 12.     Calculate the most similar pair  $P_s$  in each set of  $\mathcal{S}$   
Step 13.     Calculate  $acc_{n+i}$  based on  $P_{Dr}$  and  $P_s$   
Step 14.      $w_i++$   
Step 15.     **end for**  
Step 16.      $acc_{max} = \max(acc_1, acc_2, \dots, acc_n, acc_{n+1}, \dots, acc_{2n})$ , and let  $\alpha$  be the number of maximum values.  
Step 17.     **for**  $i$  in range( $1, 2n$ )  
Step 18.     **if**  $acc_i == acc_{max}$  &&  $i \leq n$   
Step 19.     |  $w_i += 1/\alpha$   
Step 20.     **else if**  $acc_i == acc_{max}$  &&  $n + 1 \leq i \leq 2n$   
Step 21.     |  $w_{i-n} -= 1/\alpha$   
Step 22.     **end if**  
Step 23.     **end for**  
Step 24. **until**  $acc_{max}$  keeps unchanged for a defined number iterations  
Step 25. **return**  $\mathbf{w}$

\* the source code is available at <https://github.com/marlonli/PatientCohortsAnalysis>.

algorithm gradually improves the weights and accepts suboptimal solutions, as finding the optimal solution is computationally difficult.

### B. Patient Cohorts Discovery

To discover patient cohorts, we clustered the patients into cohorts with some clinical meaning. The patients being clustered into the same cohort must share similar relevant attributes, so we used the previously learned weighted attribute distance as the similarity measures during clustering.

Numerous clustering algorithms were developed for all kinds of datasets and problems. Some clustering algorithms were specifically designed for certain data distributions (e.g., EM on Gaussian-distributed data and DBSCAN on noisy data [14]). In our study, the patient context attributes can be heterogeneous, with categorical, binary, and numerical types. Hence, to achieve the best generation of our framework, we chose two commonly used clustering algorithms: k-means clustering (centroid-based) [13][14] and hierarchical clustering (connectivity-based) [14].

In addition, selecting the number of clusters is a difficult and well-known problem. Hierarchical clustering itself is widely used to intuitively decide the optimal number of clusters by visualizing the resulting dendrogram. Another widely used method is silhouette analysis [11]. The silhouette value is a measure of how similar a data point is to its own cluster (cohesion) compared to other clusters (separation). The silhouette score ranges from -1 to +1, where a high value usually indicates a better clustering configuration. We used both methods in our study.

TABLE II. CONTEXT ATTRIBUTES (1ST COLUMN), A SET OF THREE PATIENTS (2ND-4TH COLUMNS), AND THE WEIGHTS LEARNT FROM ALG.1.

Attributes	Patient A	Patient B	Patient C	...	Weights Learnt
<sup>1</sup> AGE Group	0	0	2	...	0
<sup>1</sup> Gender (male = 1)	1	1	1	...	0
<sup>2</sup> Transfer	1	1	0	...	0
<sup>2</sup> Stat	0	0	1	...	0
<sup>2</sup> Attending	0	0	0	...	0.14
<sup>3</sup> Blunt	1	1	1	...	1
<sup>3</sup> Penetrating	0	0	0	...	1.31
<sup>3</sup> Animal Bite	0	0	0	...	0.81
<sup>3</sup> Burn	0	0	0	...	0
<sup>3</sup> No Injury	0	0	0	...	0.81
<sup>2</sup> Non-critical admission	1	1	0	...	1
<sup>2</sup> Critical Admission	0	0	0	...	0
<sup>2</sup> Discharged	0	0	1	...	0
<sup>2</sup> ETA Now	0	0	1	...	0
<sup>2</sup> Weekend	1	0	0	...	0
<sup>2</sup> Intubation	0	0	0	...	0.14
<sup>2</sup> Daytime	0	0	1	...	0
<sup>3</sup> GCS>13	1	1	1	...	0.14
<sup>3</sup> ISS Group	0	0	0	...	0
<sup>3</sup> AIS HEAD NECK	2	0	0	...	0
<sup>3</sup> AIS FACE	0	0	0	...	0
<sup>3</sup> AIS CHEST	0	0	0	...	2.39
<sup>3</sup> AIS ABD PELVIC	0	0	0	...	0
<sup>3</sup> AIS EXTREMITIES	0	0	0	...	0.25
<sup>3</sup> AIS EXTERNAL	0	1	1	...	0
<sup>3</sup> Maximum AIS	2	1	1	...	0

<sup>1</sup>Patient demographics  
<sup>2</sup>Trauma attributes  
<sup>3</sup>Injury information

### C. Workflow Discovery And Sequential Pattern Mining

To discover the treatment patterns and procedures of different patient cohorts, we used two different techniques: workflow discovery and sequential pattern mining. Workflow discovery [5], the core component of process mining, aims to automatically extract workflow models from activity logs. In this study, we used Disco (<https://fluxicon.com/disco/>), a process mining tool based on a fuzzy workflow mining algorithm [17]. However, for complex processes like ours, these workflow discovery algorithms often generate spaghetti-like models [5] which are difficult to interpret. Searching for differences between several treatment workflow models can be even more challenging. The sequential pattern approach can help address this limitation. Treatment patterns can be discovered from activity traces using sequential pattern mining algorithms. Although numerous sequential patterns may be found, the significance of the patterns can be evaluated using statistical methods. Only statistically significant patterns need manual analysis. In our implementation, we used SPADE [18], an efficient algorithm for mining frequent sequential patterns.

### D. Statistical Analytics

Statistical analyses were used to study the associations between the treatment patterns and patient cohorts. Differences in the number of treatment patterns between patient cohorts were compared using Student's t-test [15]. Student's t-test determines whether the means of two sets of

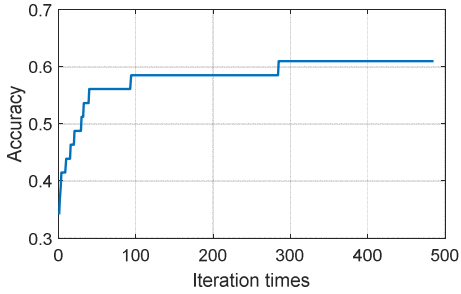


Figure 1. Attribute weight learning process. The accuracy increases across iterations.

data (dataset size can be small) are significantly different. Two different kinds of comparisons can be performed for patient cohort analysis: comparison between cohort pairs (one-vs.-one), or one cohort with the rest of the cohorts (one-vs.-rest). We set our statistical significance level at  $p < 0.05$  [15].

#### IV. EMPIRICAL EXPERIMENTAL RESULTS

Our experimental results involve three aspects: the performance of the attribute learning algorithm, the patient cohorts we discovered, and the resuscitation patterns with their significance test results.

##### A. Attribute Weights

The initial accuracy (Eq. 2) using unit weights (Figure 1) was 0.34 before applying the weight learning algorithm. The accuracy is monotonically increasing (Alg.1), rising quickly in the first 50 iterations before slowing down and converging at 0.61. Our results (Table II) show that the injury features (injury mechanism, injury area, and severity score) are more important than patient demographics (age, gender, etc.) and resuscitation attributes (i.e., stat, attending, daytime, etc.). Most patient demographics and resuscitation features have zero weights. Two resuscitation attributes with non-zero weights were “attending” and “intubation”. “attending” is a level of the categorical variable “Trauma Activation Level” (Table I (B)), so is “stat” and “transfer”. Both “stat” and “attending” level activations represent the patient arriving from the scene of injury, with stat being a standard acuity level activation and attending being the highest acuity level activation. The “intubation” is an attribute indicating whether the patients were intubated prehospital. The patients who could not maintain their airway were intubated. Within injury features, the injury mechanism attributes (penetrating, blunt, etc.) had the most non-zero weights. “AIS\_CHEST” (severity level of an injury to the chest) had the highest weight among the injured body regions.

##### B. Patient Cohorts

The silhouette analysis suggested the number of clusters to be either four or eight (two peaks in Figure 2 (c)). From the dendrogram (Figure 2 (d)), we can easily identify four clusters. We thus decided on four clusters. Although the silhouette score is higher with eight clusters, our result shows the

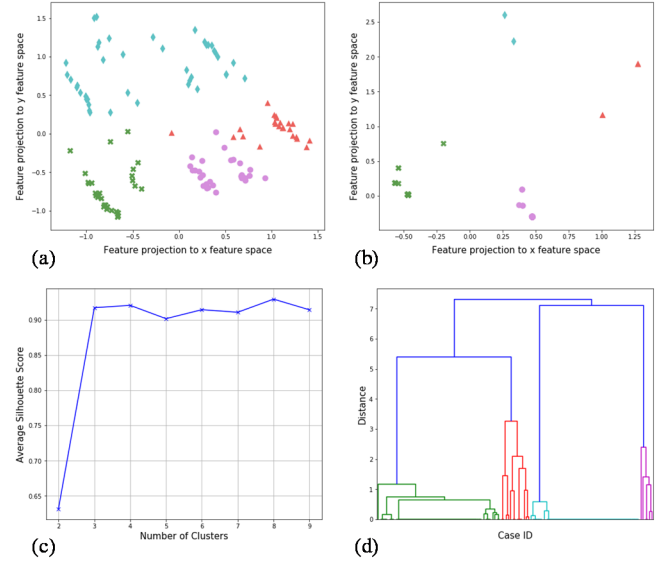


Figure 2. (a) K-means clustering results of 123 patients without unit weights. The number of clusters ( $k$ ) was set to 4. Dots of the same color are from the same cluster. (b) K-means clustering results ( $k = 4$ ) of 123 patients using weighted attributes. (c) The value of average silhouette (y-axis) changes as the number of clusters (x-axis). (d) The dendrogram plot of the hierarchical clustering results.

additional four clusters were partitioned from the two smaller clusters of the four clusters. Some of them were too small (few data points), making them a better fit for specific case studies rather than being included in a cohort analysis.

Our result also showed that k-means and hierarchical clustering returned the exact same four clusters (Figure 2 (b)(d)). The k-means clustering result is visualized after dimension reduction with principal component analysis [16] (Figure 2 (b)). We also ran the clustering on the same dataset with unit attribute weights, generating very different results (Figure 2 (a) vs. (b)). With unit attribute weights, four similar-sized clusters were formed. No clear boundary can be noted between the clusters. The distribution of data points (a) is much sparser than that of (b) because all attributes were taken in the similarity measurement. As it is rare that multiple patients have very similar context attributes, the data points in (a) cannot be distributed as densely as the clusters in (b), where only important attributes were taken into account (Table II). Four patient cohorts (from cohort 0 to cohort 3) include 55, 13, 49, and 6 patients respectively. To better understand the characteristics of each patient cohort, we used radar charts (Figure 3) to help visualize attribute distribution within each patient cohort. We also calculate the significance of each attribute in its own cohort versus the rest of the cohorts (Table III). Of the 26 attributes, we visualize only the ones with non-zero weight (Table II). Both cohort 0 and cohort 2 include patients with blunt injuries and GCS over 13. The difference is in whether the patients were critical admitted. Cohort 1 and 3 are two smaller cohorts. Their distributions are very different from the rest. Cohort 1 includes patients of the

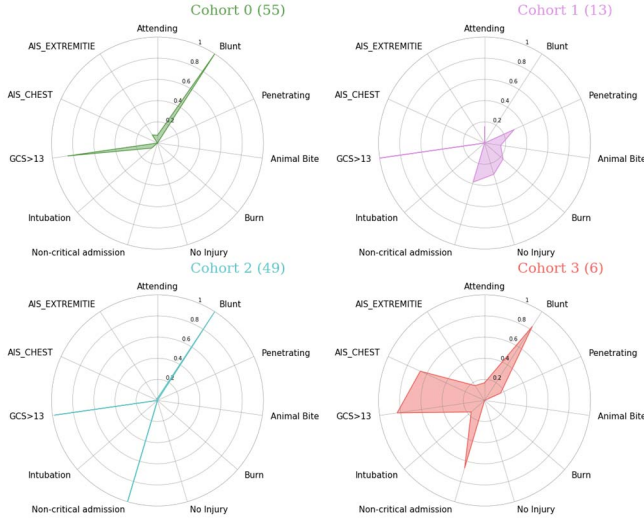


Figure 3. Radar charts visualizing the characteristics of the patient cohorts. Each radar chart represents a patient cohort. Each attribute is depicted by the node on the spoke. The number of patients in each cohort is shown in the parenthesis, e.g., cohort 0 has 55 patients.

remaining injury mechanisms except blunt. Only cohort 3 includes patients with chest injury .

### C. Resuscitation Workflow and Patterns

With patient cohorts identified, we performed workflow mining and sequential pattern mining on each cohort. The complete workflow models are spaghetti-like because of the large number of non-zero transitions. To obtain descriptive and interpretable workflow models, we applied two model simplification methods. First, we focused our study on a specific medical phase at a time. A medical phase is a part of the complete trauma resuscitation process, e.g., airway assessment phase (checking patient’s airway), disability assessment phase (assessing patient’s disability level), head assessment phase (assessing injuries on patient’s head), etc. In this paper, we use the head assessment phase as a case study for workflow and sequential pattern analysis. Second, we pruned the workflow model by only preserving the most dominant incoming and outgoing transitions for each node. This method omitted many rare insignificant transitions.

TABLE III. P-VALUES OF EACH ATTRIBUTE OF A COHORT COMPARED TO OTHER COHORTS.

Attributes Name	Cohort 0	Cohort 1	Cohort 2	Cohort 3
Attending	0.758	0.172	0.104	0.305
Blunt	0.000	0.000	0.001	0.679
Penetrating	0.040	0.000	0.064	0.111
Animal Bite	0.203	0.000	0.249	0.749
No Injury	0.068	0.000	0.100	0.648
Non-critical admission	0.000	0.511	0.000	0.330
Intubation	0.107	0.437	0.064	0.111
GCS>13	0.019	0.260	0.045	0.437
AIS_CHEST	0.042	0.375	0.034	0.000
AIS_EXTREMITIES	0.756	0.529	0.208	0.214

Our workflow results (Figure 4) on the head phase showed high similarity to the four head assessment workflows. All of them follow a similar sequential order “head (H) -> face (F) -> (nose (N) -> mouth (M)) || (eye (EY)) -> ear (EAR)”. Two differences can be seen between the workflows of cohort 0 and cohort 2. First, the occurrence of “visual inspection-head” is more frequent in cohort 0 than that in cohort 2. On the other hand, “left/right visual inspection-eye” are much more frequent in cohort 2. The medical explanation is that activities “left/right visual inspection-eye” can be optional, because there is another pair of activities “right/left pupil check” in the disability assessment phase (prior to the head assessment phase) to evaluate the patient’s neurological level. The pupil examination requires a light source to be used to assess pupil response. This exam is more thorough than only performing an unaided visual examination of the eye. In addition, because most patients in cohort 0 had a critical admission, these patients tended to have more severe conditions and the medical team tended to omit a second eye exam in favor of minimizing the time it took to transport these patients to the intensive care unit.

By performing sequential mining algorithms on patient cohorts 0 and 2, we discovered 39,784 sequential patterns in total. 178 sequential patterns were statistically significant (a few are shown in Table IV). For example, “visual inspection-head” is found to occur on average 2.87 times in cohort 0 versus 2.06 times in cohort 2 ( $p$ -value = 0.026). Similarly, “visual inspection-back” is found to occur on average 1.77 times in cohort 0 versus 1.27 in cohort 2 ( $p$ -value = 0.001). The potential medical explanation is that the patients in cohort

TABLE IV. SIGNIFICANT RESUSCITATION PATTERNS DISCOVERED FROM COHORT 0 AND COHORT 2

Resuscitation Patterns	Occur. Frequency	Raw Count (Average)	p-value
Visual inspection-H	96.43%   97.96%	155 (2.87)   99 (2.06)	0.026
Visual inspection-RUE	89.29%   93.88%	120 (2.4)   76 (1.65)	0.046
Visual inspection-BK	94.64%   100.00%	94 (1.77)   62 (1.27)	0.001
Visual inspection-H → ⋯ → R otoscopy-Ear	42.59%   10.20%	23   5	0.000
Visual inspection-H → ⋯ → Palpation-RLE	29.63%   4.08%	16   2	0.000
Log roll-BK → ⋯ → Visual inspection-BK, T-spine-BK	24.07%   2.04%	13   1	0.001
Visual inspection-M → Visual inspection-N	33.33%   8.16%	18   4	0.002
Visual inspection-BK, T-spine-BK → ⋯ → L-spine-BK	22.22%   2.04%	12   1	0.002

“→”-direct sequence; “→⋯→”-intervening tasks allowed



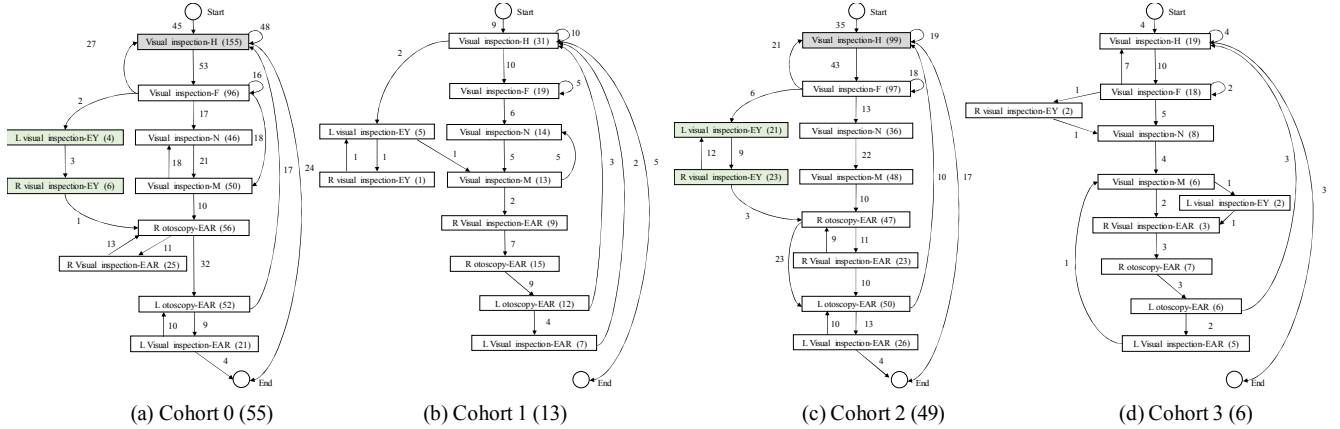


Figure 4. Workflow models discovered from patient cohorts. The major differences between cohort 0 and cohort 2 are highlighted in the figure. Each node includes an activity type and the count of its occurrences. The transition represents the sequential order of activities and the numbers on the transitions represent the count of such sequential pattern.

0 were of critical admission types, indicating they may have more severe injuries than patients of cohort 2. Hence, it is more likely that after the medical team members perform rapid evaluation, they need to confirm the initial findings by reassessing the patient.

## V. RELATED WORK

Patient cohort analysis is popular in the medical domain [1]-[4]. Traditional patient cohort analysis usually involves statistically testing a hypothesis across patient cohorts defined by targeted attributes. For example, Nelson et al [3] first proposed obese trauma patients had an increased risk of hypovolemic shock. They then defined three patient cohorts based on attribute body mass index (BMI), and showed that obese patients had a statistically-significant higher mortality rate due to hemorrhagic shock on admission. These studies are limited to only focus on the findings and cohorts medical experts are interested in, and are very likely to miss cohorts and findings they are not already familiar with.

In addition, limited research focuses on both cohort analysis and process analysis. Two recent medical recommender systems incorporated both analyses. One treatment recommender system for cerebral infarction was recently presented based on electronic medical records (EMR) [20]. Typical treatment regimens and patient cohorts were extracted, and then recommendations were provided based on the most effective regimen for a given cohort. Another process-related work [21] adopted a similar framework. It clustered traces of process executions and extracted a prototypical execution of each cluster. A regression model was then trained to find the associations between process executions and patient attributes.

Recently, data visual analytics tools for cohort and process analysis have been developed. CoCo [19] can be used to find differences between two groups (cohorts) of process traces and to highlight their significant distinguishing features (e.g., activity order, frequency, and duration). CAVA was designed to help domain experts work more independently and more quickly when performing retrospective cohort studies and

temporal visualization of medical pathways [22]. VIT-PLA [23] was designed to compute and visualize the typical workflow procedures for each patient cohorts. Similar to Yang’s work [21], VIT-PLA finds patient cohorts according to the similarity of treatment procedures. In contrast to these studies, our work identifies patient cohorts based on context attributes and analyzes the treatment patterns across these patient cohorts.

## VI. CONCLUSION

We introduced a framework for analyzing associations between treatment procedures and patient cohorts. The framework works by learning weights of context attributes by best-first search, deciding patient cohorts with clustering algorithms, discovering treatment patterns with process mining techniques, and analyzing the cohort-vs.-procedure through statistical analysis. We deployed and evaluated this framework on a complex real-world medical process, the trauma resuscitation. We uncovered four different patient cohorts from this dataset and discovered statistically significant treatment patterns across different patient cohorts. Our study provides evidence that the context attributes can affect treatment patterns (regimen) in trauma resuscitation. In the future, AI-based data-driven treatment recommender or decision support systems may be implemented in real-world medical process settings (e.g., surgical process, clinical process and trauma resuscitation process). A cohort-based system may have advantages over a universal system in this setting, as it can provide more accurate and personalized recommendations and decision support. On the other hand, this framework also suffers from a limitation that the data-driven clustering approach may lack consistency when different clustering algorithms or similarity metrics are used.

## ACKNOWLEDGMENT

This paper is based on research supported by National Institutes of Health under grant number 1R01LM011834-01A1.

## REFERENCES

- [1] Kleinman, Nathan, et al. "Cohort analysis assessing medical and nonmedical cost associated with obesity in the workplace." *Journal of occupational and environmental medicine* 56.2 (2014): 161-170.
- [2] Delgado, João, et al. "Outcomes of treated hypertension at age 80 and older: Cohort analysis of 79,376 individuals." *Journal of the American Geriatrics Society* 65.5 (2017): 995-1003.
- [3] Nelson, Jana, et al. "Obese trauma patients are at increased risk of early hypovolemic shock: a retrospective cohort analysis of 1,084 severely injured patients." *Critical Care* 16.3 (2012): R77.
- [4] Kelleher, Deirdre C., et al. "Effect of a checklist on advanced trauma life support workflow deviations during trauma resuscitations without pre-arrival notification." *Journal of the American College of Surgeons* 218.3 (2014): 459-466.
- [5] Van Der Aalst, Wil. *Process mining: discovery, conformance and enhancement of business processes*. Springer Science & Business Media, 2011.
- [6] Yang, Sen, et al. "Medical Workflow Modeling Using Alignment-Guided State-Splitting HMM." *Healthcare Informatics (ICHI), 2017 IEEE International Conference on*. IEEE, 2017.
- [7] Rozinat, Anne, and Wil MP Van der Aalst. "Conformance checking of processes based on monitoring real behavior." *Information Systems* 33.1 (2008): 64-95.
- [8] Christov, Stefan C., George S. Avrunin, and Lori A. Clarke. "Online deviation detection for medical processes." *AMIA Annual Symposium Proceedings*. Vol. 2014. American Medical Informatics Association, 2014.
- [9] Carter EA, Waterhouse JL, Kovler ML, Fritzeen J, Burd RS. Adherence to ATLS primary and secondary surveys during pediatric trauma resuscitation. *Resuscitation*. 2013 Jan;84(1):66-71. PubMed PMID: 22781213.
- [10] Hand, David J., Heikki Mannila, and Padhraic Smyth. *Principles of data mining (adaptive computation and machine learning)*. Cambridge, MA: MIT press, 2001.
- [11] Liu, Yanchi, et al. "Understanding of internal clustering validation measures." *Data Mining (ICDM), 2010 IEEE 10th International Conference on*. IEEE, 2010.
- [12] Cormen, Thomas H. *Introduction to algorithms*. MIT press, 2009.
- [13] MacQueen, James. "Some methods for classification and analysis of multivariate observations." *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*. Vol. 1. No. 14. 1967.
- [14] Rokach, Lior, and Oded Maimon. *Clustering methods*. Data mining and knowledge discovery handbook. Springer, Boston, MA, 2005. 321-352.
- [15] De Winter, Joost CF. "Using the Student's t-test with extremely small sample sizes." *Practical Assessment, Research & Evaluation* 18.10 (2013).
- [16] Shlens, Jonathon. "A tutorial on principal component analysis." *arXiv preprint arXiv:1404.1100* (2014).
- [17] Günther, Christian W., and Wil MP Van Der Aalst. "Fuzzy mining—adaptive process simplification based on multi-perspective metrics." *International Conference on Business Process Management*. Springer Berlin Heidelberg, 2007.
- [18] Zaki, Mohammed J. "SPADE: An efficient algorithm for mining frequent sequences." *Machine learning* 42.1-2 (2001): 31-60.
- [19] Malik, Sana, et al. "Cohort comparison of event sequences with balanced integration of visual analytics and statistics." *Proceedings of the 20th International Conference on Intelligent User Interfaces*. ACM, 2015.
- [20] Sun, Leilei, et al. "Data-driven Automatic Treatment Regimen Development and Recommendation." *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2016.
- [21] Yang, Sen, et al. "A Data-driven Process Recommender Framework." *Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. ACM, 2017.
- [22] Zhang, Zhiyuan, David Gotz, and Adam Perer. "Iterative cohort analysis and exploration." *Information Visualization* 14.4 (2015): 289-307.
- [23] Yang, Sen, et al. "VIT-PLA: Visual Interactive Tool for Process Log Analysis." *KDD 2016 Workshop on Interactive Data Exploration and Analytics (IDEA 2016)*.