Analyzing the Trajectories of Patients with Sepsis using Process Mining

Felix Mannhardt[1] ⊠, Daan Blinde[2]

[1] Eindhoven University of Technology, Eindhoven, The Netherlands
[2] Blijnder, The Netherlands
f.mannhardt@tue.nl,daan@blijnder.nl

Abstract. Process mining techniques analyze processes based on event data. We analyzed the trajectories of patients in a Dutch hospital from their registration in the emergency room until their discharge. We considered a sample of 1050 patients with symptoms of a sepsis condition, which is a life-threatening condition. We extracted an event log that includes events on activities in the emergency room, admission to hospital wards, and discharge. The event log was enriched with data from laboratory tests and triage checklists. We try to automatically discover a process model of the patient trajectories, we check conformance to medical guidelines for sepsis patients, and visualize the flow of patients on a de-jure process model. The lessons-learned from this analysis are: (1) process mining can be used to clarify the patient flow in a hospital; (2) process mining can be used to check the daily clinical practice against medical guidelines; (3) process discovery methods may return unsuitable models that are difficult to understand for stakeholders; and (4) process mining is an iterative process, e.g., data quality issues are often discovered and need to be addressed.

Keywords: Process Mining · Medical Guidelines · Patient Trajectories

1 Introduction

Process models are used by organizations to document, specify, and analyze their processes. A process model describes the expected behavior of a process. Any set of related activities that are executed in a repeatable manner and with a defined goal can be seen as process. For example, in a health care context the trajectory of a patient from arrival in the emergency ward to the admission to a hospital ward and up to the discharge can be seen as a process. Often the execution of such a processes is supported by information systems. For example, the hospital may record medical information such as symptoms, the condition upon arrival of the patient, and the results of blood tests. Moreover also logistical information are recorded such as the movement of patients between wards and different types of discharge.

Process mining uses data recorded by such systems to analyze the actual execution of processes. Common goals of process mining are: *process discovery*,

the discovery of accurate process models that help in understanding the process execution and *conformance checking*, the diagnosis of problems in the process execution by comparing the real execution (*as-is process*) to a process model that defines the desired process execution (*to-be process*). Process mining has been previously used in the health care context [1,2,3]. However, it is often stated that the flexibility of the work in a hospital makes direct application of process mining methods difficult [2].

We used state-of-the-art process mining methods to analyze the trajectories of 1050 patient that are admitted to the emergency ward of a Dutch hospital[1] because they display symptoms of a sepsis. Sepsis is a life-threatening condition that requires immediate treatment [4]. The goals of the process mining project were:

1. to get insights in the patient trajectories,
2. to validate the compliance with medical guidelines for sepsis treatment, and
3. to evaluate the use of process mining methods in this context.

2   Context and Data Collection

We conducted the process mining project in a regional hospital in The Netherlands. The hospital has about 700 beds at several locations and is visited by about 50,000 patients per year. The scope of our project was on the patient trajectories of patients that are admitted to the emergency ward. We focused on a *specific sub group of patients* to avoid a common problem in process mining for health care processes: the inherent complexity and flexibility of health care processes [2]. Scoping our analysis to a single group of patients, for which a specific treatment is to be expected, avoids some complexity.
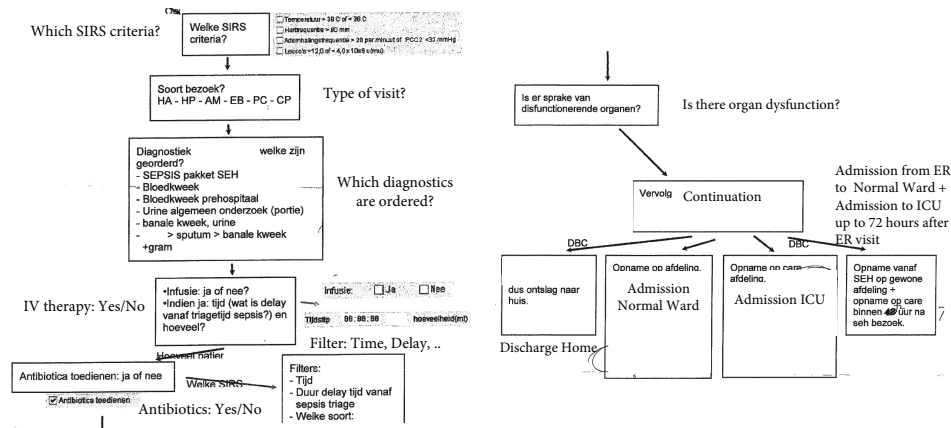

Fig. 1. Patient flow and questions as described by the process stakeholders.

---

[1] For privacy reasons we do not disclose the name of the hospital.

Figure 1 shows an illustration of the patient flow that was created by process stakeholders from the emergency department. This document served as starting point for the process mining analysis. It helped to clarify the assumptions and perspective on the process of nurses and doctors in the emergency ward. We identified several questions:

1. are particular medical guidelines for the treatment of sepsis patients followed:
   − patients should be administered antibiotics within one hour,
   − lactic acid measurements should be done within three hours;
2. visualize and investigate the following specific trajectories:
   − discharge without admission,
   − admission to the normal care,
   − admission to the intensive care,
   − admission to the normal care and transfer to intensive care;
3. investigate the trajectory of patients that return within 28 days.

Moreover, the document in Figure 1 facilitated the collection of event data. Based on the document, we could identify several sources for events:

− the triage document filled in for sepsis patients with information on
  • the time the triage was conducted,
  • the symptoms present (SIRS criteria for sepsis),
  • the diagnostics ordered,
  • the time infusions of liquid and antibiotics were administered;
− documents created by the laboratory for several blood tests,
− information about the further trajectory of the patient recorded by financial systems.

All documents and records are stored in different databases of the ERP system that is used by the hospital.
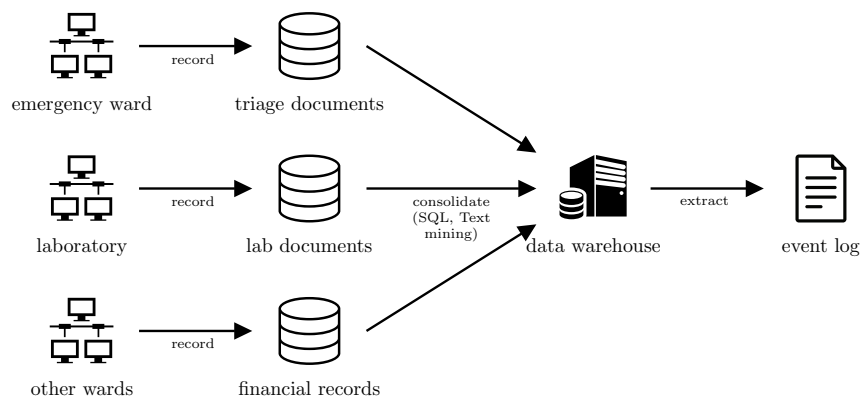

Fig. 2. Consolidation of data from several source systems to a single event log

Based on the identified sources, we collected data about several activities that are executed for this group of patients from three different systems of the hospital. The activities can be coarsely categorized into *medical activities* and *logistical activities*.

Figure 2 shows how we collected the event data and consolidated it into a single anonymized event log covering the traces that were recorded for 1050 patients over the course of 1.5 years in the hospital information systems. The resulting event log is made available for further process mining research purposes [5]. The event log contains events for 16 activities:

- 3 activities regarding the registration and triaging in the emergency ward;
- 3 activities regarding measurement of leukocytes, CRP, and lactic acid;
- 2 activities regarding admission or transfer to normal care or intensive care;
- 5 activities for variants of discharge from the hospital; and
- an activity concerned with returning patients at a later time.

3   Applied Process Mining Methods

We applied several process mining methods on the event log. We started by applying the Inductive Miner (IM) [6] as a state-of-the-art process discovery method. Figure 3 shows the returned process model. Some of the structures that we expected were discovered by the IM. For example, the process starts with patient registration (ER Registration) and ends with some variants of discharge (Release C-E). However, most of the activities that occur in between these registration and discharge activities are modeled very imprecisely. Almost all activities can be skipped and repeated. We highlighted the loop edge in Figure 3.
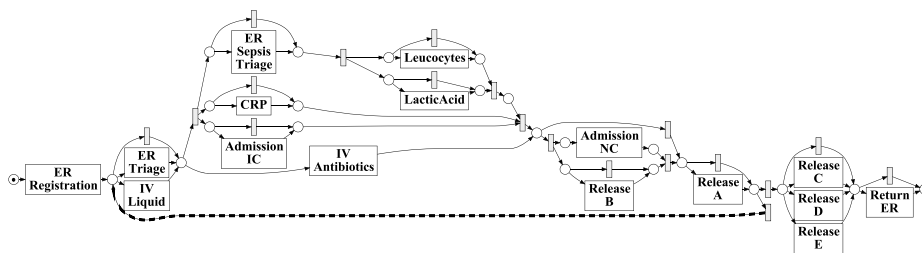

Fig. 3. Model discovered by the Inductive Miner.

Therefore, the process model returned by IM is difficult to be used for *the communication with doctors and nurses*. At a first glance it appears that administering antibiotics (IV Antibiotics) and filling in the triage document (ER Sepsis Triage) is in an exclusive choice. Yet, it is possible to repeat the entire middle part of the process, thus, both activities can occur in the same process instance. Using different parameter settings for the IM did not help. Even though we could obtained a more precise process model, unfortunately, that process model does not

allow the laboratory tests to be repeated. Moreover, it does not allow to visualize the patient trajectory: first admission to normal care and, then, a transfer to intensive care. This trajectory is important to address Question 2. Since we use process discovery to address concrete questions, we did not rely on measures such as *fitness* and *precision* to estimate the quality of the model. Instead, we determined the model quality based on whether it is suited to answer our process questions and whether it is helpful in communicating with doctors and nurses.

As an alternative to automatic process discovery, we iteratively designed a suitable process model based on domain knowledge obtained from doctors and nurses working in the emergency ward. One example for such domain knowledge are the descriptions shown in Figure 1. In each iteration, we used alignment-based conformance checking techniques to validate whether the hand-drawn process model reflects the observations in the event log. Since the iteratively designed model follows the expectations of doctors and nurses working in the emergency ward, it is can be used (1) to communicate with the stakeholders, and (2) to answer the process questions.

Figure 4 shows the iteratively designed model. It fits 98.3% of the event log, i.e., almost all of the events can be correctly replayed on the model. We used the ProM plug-in Multi-perspective Process Explorer (MPE) [7] to project the event log onto the process model and explore the process. Figure 4 also depicts the output of the MPE. The edges are scaled according to the number of cases that are projected on the edge. The Petri net transitions are colored according to the conformance level, the darker a transition the more conformance violations. We used the MPE to explore the three identified questions about the process.

Question 1: medical guidelines There are two time rules specified by the sepsis guideline:

1. between *ER Sepsis Triage* and *IV Antibiotics* should be less than 1 hour,
2. between *ER Sepsis Triage* and *LacticAcid* should be less than 3 hours.

We used Data Petri nets (DPN) to encode both time-perspective rules using three variables *timeTriage*, *timeAntibiotics*, and *timeLacticAcid*. We added guard expressions to both activities *IV Antibiotics* (timeAntibiotics' $\leq$ timeTriage + 60) and *LacticAcid* (timeLacticAcid' $\leq$ timeTriage + 180). Then, we used the multi-perspective conformance checking technique that is described in work [8] to check conformance. Rule 1 regarding administering antibiotics within one hour is sometimes violated: *IV Antibiotics* and *timeAntibiotics* in Figure 4 are colored orange. The average time between *ER Sepsis Triage* and *IV Antibiotics* is 1.7 hours and the rule is violated 58.5% of the cases. Regarding the second rule the situation is much better. Rule 2 regarding the timely measurement of lactic acid is only violated in 0.7% of the cases. The high number of violations regarding rule 1 can be explained be two factors:

– Guidelines recommend to administer antibiotics within one hours, however, acknowledged that this is not always feasible [4]. Moreover, not all patients in our event log show symptoms of a severe sepsis. Thus, the one hour rule can be considered very strict.
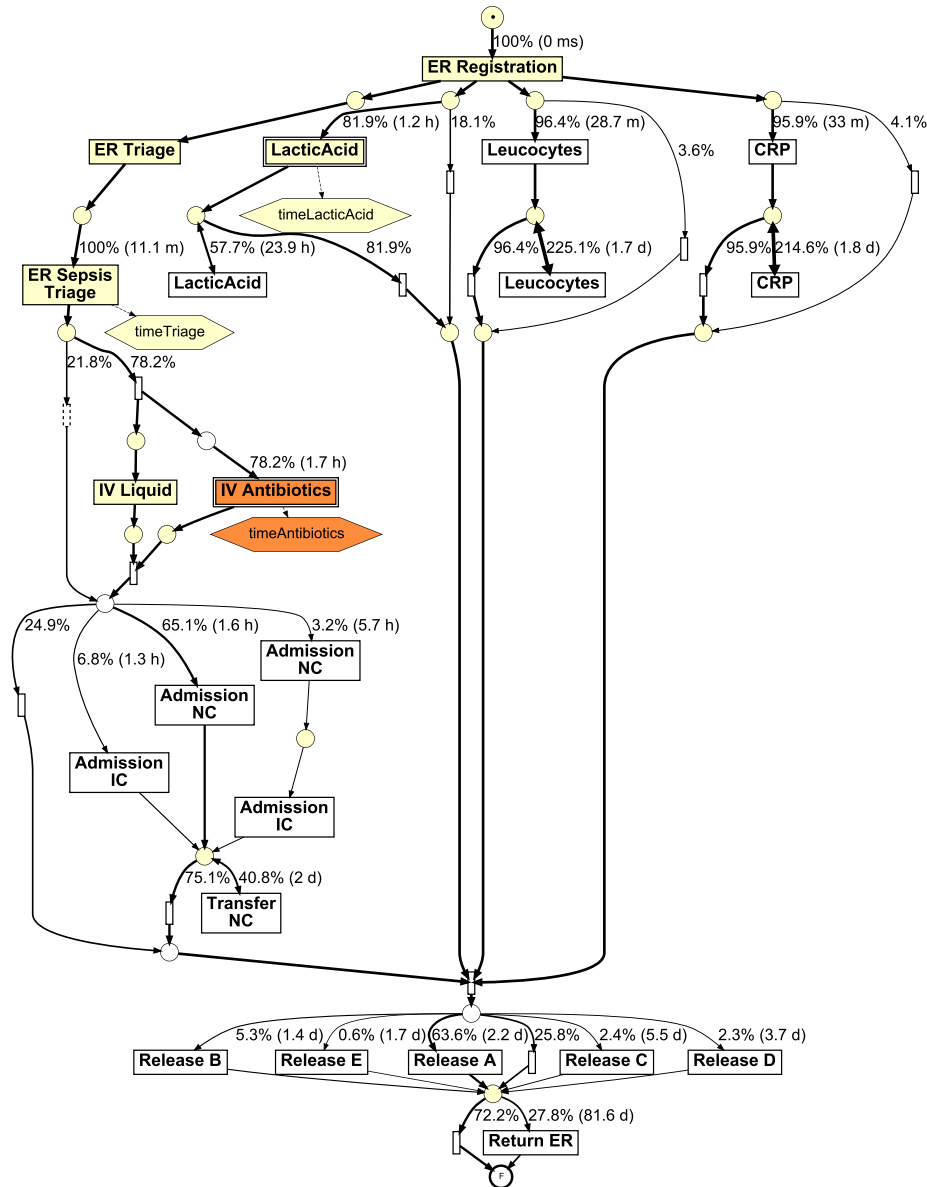
Fig. 4. Projection of the events on the hand-made process model for the sepsis process as returned by the Multi-perspective Explorer. Darker colored transition and variables correspond to conformance problems. Edges are annotated with the frequency relative to the number of traces and the average time between activities.

Fig. 5. Digital triage form that is filled out in the emergency ward.

– The data about the infusion of antibiotics is entered manually by nurses and doctors in a form as shown in Figure 5. When looking at the data, we found some cases in which the entered antibiotics timestamp is 24 hours after the triage and other cases in which the antibiotics timestamp was before the triage document was filled in. Thus, it is difficult to conclude actions from the measurement, since it might be caused by poor data quality.

Question 2: patient trajectories Using the MPE, we could visualize the four trajectories of interest. In Figure 4 they are represented as an exclusive choice in the main branch of the process on the right side. The choice regarding the admission of patients is made after filling in the general triage document (*ER Triage*), the triage document for patients with suspicion for a sepsis (*ER Sepsis Triage*), and, possibly, giving infusions of antibiotics and liquid (*IV Liquid* and *IV Antibiotics*). It can be seen that 18.1% of the patients leave the emergency ward without admission to the hospital. Most patients (70.6%) are admitted to the normal care ward (*Admission NC*), less patients (6.8%) are admitted to the intensive care ward (*Admission IC*), and a small group of patients (3.6%) is first admitted to the normal care ward and, directly afterwards, admitted to the intensive care ward. The latter group of patients is of interest since their condition got worse. Therefore, it would be beneficial to avoid patients following this trajectory.

Question 3: returning patients We tried decision mining techniques such as described in [9,10] to discover rules regarding patients that return within some time to the emergency ward. In total 27.8% of the patients return to the emergency room. On average it takes 81.6 days for a patient to return. We were only interested in patients that returned within 28 days, thus, we filtered the event log accordingly. Out of all patients 12.6% return within 28 days. Unfortunately, we did not find any decision rule based on the available data attributes of the event log, which include checkboxes marked in the triage form and the values of blood tests.

4   Lessons Learned

We now report on the lessons learned from analyzing an event log recorded for sepsis patients in a regional hospital with process mining technology. The project was successful in visualizing the trajectories of patients with a suspicion for sepsis.

In fact, a doctor found the process mining analysis to be "[a] magnificent way to clarify patient flow".

Unfortunately, little actionable results were obtained. We attribute this to the lack of a set of initial hypotheses that could lead to actionable result. We started the project with some process questions such as the compliance to the medical guidelines and patients returning within 28 days. Still, it turned out to be difficult to follow-up on the findings due to lack of data quality regarding the time when antibiotics were given and the general lack of data that could explain the returning patients. In conclusion, we found that:

1. Process mining can be used to clarify and visualize the flow of patients in a hospital. This confirms previously reported results in [1,2,3]. However, we show that when focusing on a specific group of patients and a mix of medical and logistical activities, the often cited flexibility of health care processes [1,2,3] can be avoided.
2. Often hospitals monitor their processes by looking only at quality indicators (e.g., length of the hospitalization, percentage of on-time surgeries). Process mining can be used to check the conformance to medical guidelines by hospitals in the context of treatment and patient logistics processes. Medical guidelines [11] can be encoded as part of the process model and the whole process (i.e. pathways, outcomes, waiting times, and guidelines) can be monitored with an alignment technique [8]. Similar results have been presented for a declarative modeling notation in [3]. We show that procedural notation are also suitable when focusing on a specific process.
3. Process discovery methods may return models of the observed behavior that are unsuitable to communicate with stakeholders and answer questions. The discovered process model might correctly represent the observed behavior. However, its structure may not be recognizable to the people working in the process. Iteratively creating a hand-made model proved to be useful and feasible in this case.
4. Obtaining the data and analyzing the process is an iterative process. Data quality issues and further questions may arise only after some initial data has been collected. Regular feedback from process stakeholders to validate assumptions is important. Iterative analysis and data enrichment has also been proposed in process mining methodologies such as [12]. In our case we also extracted more data from the source systems after an initial analysis iteration.

References

1. Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B.: Process Mining in Healthcare. Springer (2015)

2. Rebuge, Á., Ferreira, D.R.: Business process analysis in healthcare environments: A methodology based on process mining. Information Systems **37**(2) (2012) 99–116
3. Rovani, M., Maggi, F.M., de Leoni, M., van der Aalst, W.M.: Declarative process mining in healthcare. Expert Syst. Appl. **42**(23) (2015) 9236 – 9251
4. Rhodes, A., et al.: Surviving sepsis campaign: International guidelines for management of sepsis and septic shock: 2016. Intensive Care Med. **43**(3) (2017) 304–377
5. Mannhardt, F.: Sepsis Cases - Event Log. Eindhoven University of Technology. Dataset. (2016) doi:10.4121/uuid:915d2bfb-7e84-49ad-a286-dc35f063a460.
6. Leemans, S.J.J., Fahland, D., van der Aalst, W.M.P.: Discovering block-structured process models from event logs - A constructive approach. In: Petri Nets. Volume 7927 of LNCS., Springer (2013) 311–329
7. Mannhardt, F., de Leoni, M., Reijers, H.A.: The multi-perspective process explorer. In Daniel, F., Zugal, S., eds.: BPM (Demos). Volume 1418 of CEUR Workshop Proceedings., CEUR-WS.org (2015) 130–134
8. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Balanced multi-perspective checking of process conformance. Computing **98**(4) (2016) 407–437
9. de Leoni, M., van der Aalst, W.M.P.: Data-aware process mining: Discovering decisions in processes using alignments. In: SAC'13. SAC '13, New York, NY, USA, ACM (2013) 1454–1461
10. Mannhardt, F., de Leoni, M., Reijers, H.A., van der Aalst, W.M.P.: Decision mining revisited - discovering overlapping rules. In Nurcan, S., Soffer, P., Bajec, M., Eder, J., eds.: CAiSE. Volume 9694 of LNCS., Springer (2016) 377–392
11. Peleg, M., Tu, S., Bury, J., Ciccarese, P., Fox, J., Greenes, R.A., Hall, R., Johnson, P.D., Jones, N., Kumar, A., Miksch, S., Quaglini, S., Seyfang, A., Shortliffe, E.H., Stefanelli, M.: Comparing computer-interpretable guideline models: A case-study approach. J. Am. Med. Inform. Assoc. **10**(1) (jan 2003) 52–68
12. van Eck, M.L., Lu, X., Leemans, S.J.J., van der Aalst, W.M.P.: PM^2: A process mining project methodology. In: CAiSE. Volume 9097 of LNCS., Springer (2015) 297–313