Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions

R.S. Mans[1], W.M.P. van der Aalst[1], R.J.B. Vanwersch[1], A.J. Moleman[2]

[1] Department of Information Systems, Eindhoven University of Technology, P.O. Box 513, NL-5600 MB, Eindhoven, The Netherlands.
r.s.mans,w.m.p.v.d.aalst,r.j.b.vanwersch@tue.nl

[2] Academic Medical Center, University of Amsterdam, Department of Quality Assurance and Process Innovation, Amsterdam, The Netherlands.
a.j.moleman@amc.uva.nl

Abstract. In hospitals, huge amounts of data are recorded concerning the diagnosis and treatments of patients. Process mining can exploit such data and provide an accurate view on healthcare processes and show how they are *really* executed. In this paper, we describe the different types of event data found in current Hospital Information Systems (HISs). Based on this classification of available data, open problems and challenges are discussed that need to be solved in order to increase the uptake of process mining in healthcare.
Key words: Visualization, monitoring and mining healthcare processes

1 Introduction

Todays hospital information systems contain a wealth of data. Typically, these systems record information about (business) processes in the form of so-called event logs. These logs can be used as input for process mining such that process-related information can be extracted from these logs. In a healthcare context, process mining can be used to provide insights into how healthcare processes are *really executed*. People involved in these processes tend to have an ideal scenario in mind, which in reality is only one of the many scenarios possible.

Currently, process mining has been applied in many organizations (e.g. municipalities, banks, government agencies) and it is attracting a huge interest from industry too [1]. Process mining has also been used in the healthcare domain (see Section 2.2 for an overview). For these applications typically only data is taken from one system in order to solve a particular problem. For example, data is taken from a particular medical department or data is taken from an administrative system. However, an overview is missing of the application possibilities of process mining within the entire hospital. Therefore, in this paper we *investigate the data challenges that are faced when answering frequently posed questions during process mining projects in hospitals*. First, we present an overview of the type of process mining questions that are frequently posed by medical professionals. Second, we investigate which process mining data can be found in current Hospital Information Systems (HISs). Also, we investigate the characteristics of

such data and whether it allows for solving the frequently posed questions. As part of this, we present a spectrum which, based on two dimensions, provides a classification of the systems in a HIS. Finally, by means of a concrete case study it is illustrated which data challenges exist when answering typical process mining questions. Finally, open problems and challenges for applying process mining in healthcare are discussed.

The outline of this paper is as follows. In Section 2, we introduce the basics of process mining and give an overview of process mining applications to health care processes. In Section 3, we outline the questions that are typically posed by medical professionals in process mining projects. In Section 4, we describe the different types of event data found in a HIS. Afterwards, in Section 5, the case study is discussed. Finally, Section 6 concludes the paper with an overview of open problems and challenges.

2 Process Mining

In this section, we first give an introduction to process mining followed by an overview of the applications of process mining in healthcare that have been identified in literature.

2.1 Overview

Process mining is applicable to a wide range of systems. The only requirement is that the system produces *event logs*, thus recording (parts of) the actual behavior. For these event logs it is important that each event refers to a well-defined step in the process (e.g. a lab test) and is related to a particular case (e.g. a patient). Also, additional information such as the performer of the event (i.e. the doctor performing the test), the timestamp of the event, or data elements recorded along with the event (e.g. the age of the patient) may be stored. Based on these event logs, the goal of process mining is to extract process knowledge (e.g. process models) in order to discover, monitor, and improve real processes [1]. As shown in Figure 1, three types of process mining can be distinguished.

Discovery: inferring process models that are able to reproduce the observed behavior. The inferred model may be a Petri net, a BPMN model, or an EPC. For example, the discovered model may describe the typical steps taken before surgery. Note that also models describing the organizational, performance, and data perspective may be discovered.

Conformance: checking if observed behavior in the event log conforms to a given model. For example, it may be checked whether a medical guideline which states that always a lab test and an X-ray needs to be done is always followed.

Extension: projection of the information extracted from the log onto the model. For example, performance information may be projected on a discovered healthcare process in order to see for which examinations a long waiting time exists.

The ProM framework and tool set has become the de facto standard for process mining. ProM (`www.processmining.org`) is a "plug-able" environment for process mining using MXML, SA-MXML, or XES as input format.

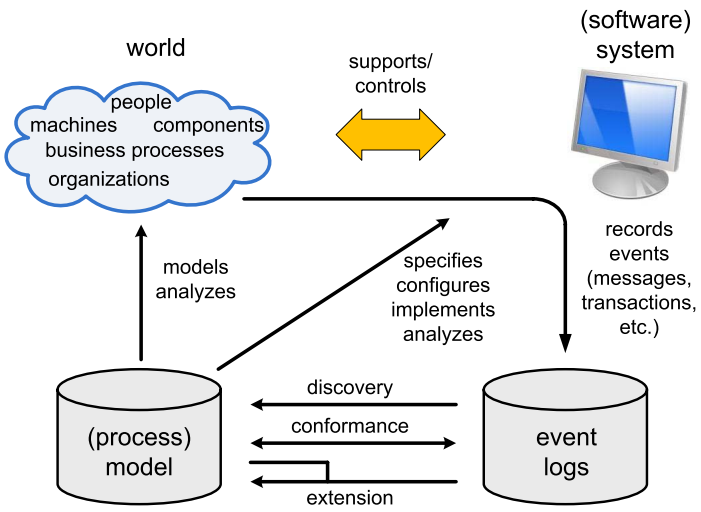


Fig. 1. Three types of process mining: (1) Discovery, (2) Conformance, and (3) Extension.

2.2 Application of Process Mining in Healthcare

There is a huge potential for process mining in healthcare as it allows hospitals to analyze and streamline their processes. However, when checking literature concerning the application of process mining in healthcare, it can be seen that it is a relatively unexplored field. In total, 12 scholarly publications have been identified in which an application of process mining in healthcare was described. That is, in [12,16] the gynaecological oncology healthcare process within an university hospital has been analyzed; in [17] several processes within an emergency department have been investigated; in [9] all Computer Tomography (CT), Magnetic Resonance Imaging (MRI), ultrasound, and X-ray appointments within a radiology workflow have been analyzed; in [21] the focus is upon the process that is followed by patients suffering from rheumatoid arthritis; in [6] the treatment of patients within an intensive care unit has been investigated; in [6, 11] process mining has been applied to different datasets for stroke patients; in [14] the activities that are performed for patients during hospitalization for breast cancer treatment are investigated; in [13] the journey through multiple wards has been discovered for inpatients; in [18] the processes for mamma care patients and diabetes foot patients has been investigated, in [2], the workflow of a laparascopic surgery has been analyzed, and finally in [5] process mining has been applied to the event logs of more than thousand X-ray machines all over the world.

These publications demonstrate that process mining can be successfully applied in the healthcare domain. Also, event log data may originate from various data sources in a hospital. For example, the data used in [12,16,18,21] originated from an administrative system within the hospital. Furthermore, data may also come from an intensive care unit [6], neurology department [11], and radiology

Type	azM	AMC	Isala	GGzE	[2]	[5]	[6]	[9]	[11]	[12]	[13]	[14]	[16]	[17]	[18]	[21]
Q1	X	X	X	X	X	X	X	X		X	X	X	X	X	X	X
Q2	X	X		X					X			X				
Q3	X	X		X									X			X
Q4	X	X							X				X	X		

Table 1. For azM, AMC, Isala klinieken, and GGzE it is indicated by a cross which type of questions have been posed by their medical professionals in process mining projects. Furthermore, for the papers that discussed an application of process mining in healthcare, it is also indicated which type of questions have been posed by them.

department [9]. Finally, process mining can also be applied to data of medical devices [2,5]. However, to date, no work exists which provides an overview of all the data in a hospital that can be used for process mining and its characteristics.

3 Questions

In this section, we give an overview of the type of questions that are frequently posed by medical professionals in process mining projects. In order to come up with a list of frequently posed questions, we have systematically analyzed the publications in Section 2.2. Furthermore, we are involved in a research project aiming at enhancing the uptake of process mining in hospitals[3]. In this project, we have analyzed datasets of academisch ziekenhuis Maastricht (azM[4]), Academisch Medisch Centrum (AMC[5]), Isala klinieken[6], and Geestelijke Gezondheidszorg Eindhoven (GGzE[7]). The questions asked by their medical professionals are included in the list. Note that of course more questions can be solved using process mining. Here, our focus is on the questions that are *frequently* posed by medical professionals. Afterwards, we investigate in Section 4 whether the process mining data that can be found in current HIS allows for answering these questions.

The list of frequently posed questions is given below. Each type of question is elaborated upon and illustrated by concrete examples. Furthermore, in Table 1 for azM, AMC, Isala klinieken, and GGzE it is indicated by a cross which type of questions have been posed by their medical professionals in process mining projects. Also, for the papers that discussed an application of process mining in healthcare, it is indicated which type of questions have been asked by them. Note that the first two questions relate to the discovery type of process mining, the third to conformance, and the last one to extension.

[3] `http://www.stw.nl/nl/content/developing-tools-understanding-healthcare-processes`
[4] `www.azm.nl`
[5] `www.amc.nl`
[6] `www.isala.nl`
[7] `www.ggze.nl`

***Q1:What are the most followed paths and what exceptional paths are followed?*:**

For the standard paths the medical specialists are mainly interested in the activities that are typically executed and the order of them. For exceptional paths, medical professionals are interested in whether this is caused by the way of working of medical specialists (inter-specialist variability) or whether this is due to a specific group of patients (e.g. medically complex patients). For example, in Isala klinieken they were interested in seeing the main process followed by urology patients in order to identify whether the process needs to be changed or not. Furthermore, for the exceptional paths they wanted to know whether these exceptions are related to specific patient characteristics or not.

***Q2:Are there differences in care paths followed by different patient groups?*:**

Here, the medical professionals are mainly interested in seeing the differences and whether subsequently process parts need to be adjusted. This comparison may not only be interesting for patient groups within a hospital but also for similar patient groups in different hospitals. For example, medical specialists in azM were interested in comparing the process of colorectal cancer patients and Hepato-Pancreato-Biliary (HPB) patients as for the first group of patients measures have been taken to optimize the process and similar measures have not been taken for the latter group. Another example can be found in Isala klinieken that wanted to compare their urology healthcare process with the same process of another hospital in order to see whether they can do things differently in their process (e.g. skipping activities).

***Q3:Do we comply with internal and external guidelines?*:**

For certain patient groups, standards are defined either by external bodies (e.g. government) or internally within the hospital itself (e.g. the medical specialists themself). An example of an external guideline is that in the Netherlands, for cancer patients, a standard is defined which indicates that for 80% of the patients, the start of the clinical treatment of the patient should happen within 5 weeks of the first visit [19]. An example of an internal guideline is that in azM, a care pathway has been defined by the medical specialists about the treatment of colorectal cancer patients and the time period in which certain activities need to be completed.

***Q4:Where are the bottlenecks in the process?*:**

One of the main motivations for using process mining is that throughput times for treating patients need to be minimized. For example, at the surgery department of AMC they had the impression that quite a long time passes before the patient is again seen on the outpatient clinic. Therefore, they were interested in seeing the medical departments for which long waiting times exist.

Obviously, all frequently posed questions are related to learning about how the current process is executed and which process areas need improvement.

4 Process Mining Data Spectrum

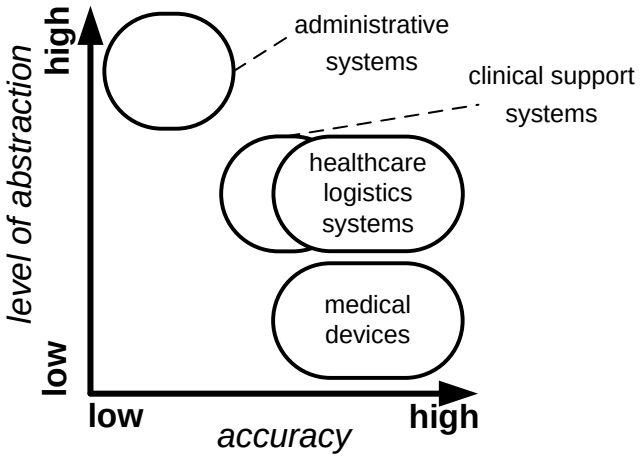
In order to be able to answer the frequently posed questions that are given in Section 3, data needs to be available about the process steps, their timing, and for which patient each step is performed. This data is typically available within a HIS. A HIS is the socio-technical subsystem of a hospital, which comprises all information processing as well as the associated human or technical actors in their respective information processing roles [8]. In this section, a spectrum is discussed which classifies the typical event data found in such systems. We do not aim to provide a full classification of all systems in a HIS. We are only interested in these systems that contain the basic process mining information (i.e. each event refers to a well-defined step in the process, is related to a particular case, and has a timestamp). As a consequence, we only consider the *information processing tools* of a HIS. Additionally, we only focus on the support of *organizational processes* within a HIS. Organizational healthcare processes capture the organizational knowledge which is necessary to coordinate collaborating healthcare professionals and organizational units (e.g. reporting of results or preparations for surgery) [10].

The spectrum that will be presented in this section is based on discussions with HIS professionals from two Dutch hospitals and the HIS classifications discussed in [8, 15]. In this way, we believe that the classification is generally applicable.

As shown in Figure 2, the spectrum distinguishes four systems and is based on two dimensions. The first dimension concerns the *level of abstraction of the events*. The level of abstraction is *average* if the event refers to an individual task, *high* if the event refers to a collection of tasks (e.g. a (sub)process), and *low* if the event refers to movements within a task (e.g. configuring a CT-scanner). The second dimension concerns the *accuracy of the timestamp* of the events. This dimension is divided into three different subdimensions. The *granularity* subdimension refers to the granularity of the timestamp and is high in case of milliseconds granularity, average in case of hour/minute granularity and low in case of day granularity. The *directness of registration* subdimension is high if timestamps are directly registered (e.g. automatically), and low if they are registered later (e.g. manually). The *correctness* subdimension is high if the timestamp is logged correctly given the chosen level of granularity and low if the timestamp is logged incorrectly given the chosen level of granularity.

The spectrum distinguishes the following systems:

Administrative systems *take care of the administration and billing of all accountable services.* The primary purpose of these systems is the registration of services that have been delivered to patients (e.g. treatments and examinations). For example, by Dutch law all hospitals need to record the diagnosis and treatment steps at the level of individual patients in order to receive payment. Therefore, it is sufficient to know that services have been delivered to patients. As a results, they may be entered manually into the system and only a day timestamp needs to be assigned to them. So, both the directness and the granularity of timestamps is low. The correctness subdimension is average as the



a) Visualization of the spectrum

	Level of abstraction	Accuracy	Granularity	Directness	Correctness
Administrative systems	High	Low	Low	Low	Average
Clinical support systems	Average	Average	Average	Low	High
Healthcare logistics systems	Average	High/Average	Average/Low	High	High
Medical devices	Low	High/Average	High	High	Average

b) For each type of system, it is indicated which value is given for the 'level of abstraction' and 'accuracy' dimension. Also, for the 'accuracy' dimension it is indicated which value is given for each subdimension.

Fig. 2. Process Mining data spectrum.

logging of timestamps may be incorrect. For example, in one hospital the wrong timestamp was recorded for one diagnostic test due to wrong batch recording,. In conclusion, the level of abstraction of events is high and the level of accuracy of timestamps is low.

Clinical support systems *involve systems of departments having such specialized needs that they require special information systems.* The purpose of these systems is to support the clinical work at a department (e.g. pathology or intensive care). As such it is required to register at a task-based level what has been performed for patients. Due to the repetitive nature of some tasks (e.g. measuring blood pressure on an intensive care unit), it is necessary to know in which hour or minute these tasks have taken place. So, the granularity subdimension of timestamps is average. Typically, these tasks are entered manually into the system. Consequently, the directness of the associated timestamp is low. Generally, these systems do not face any logging issues and hence the correctness of the timestamp of events is high. In sum, the level of abstraction of events and the level of accuracy of timestamps is average.

Healthcare logistics systems *support the logistics of operational processes.*

The primary purpose of these systems is that appointments can be made for patients and that services from medical departments can be requested, i.e. order entry and order communication. As such, events typically refer to tasks that are performed (e.g. making an appointment, filling in and sending an order). As events typically refer to tasks that are performed by people that are using the system itself, both the directness and correctness of timestamps is high. The granularity subdimension of timestamps typically varies between average and low. In conclusion, the level of abstraction of events is average and the level of accuracy of the timestamp ranges from high to average.
Medical devices *involve the systems belonging to devices that are used by medical professionals.* The goal is to collect detailed information that is useful for the manufacturer of the medical device. For example, low-level information may be recorded concerning user commands, run-time information, and errors that occur (e.g. for an X-ray machine user commands such as moving a table or capturing a single image are stored). As such, small pieces of work which are part of a task are recorded but also state information of medical devices is recorded. This information is recorded automatically. Consequently, the directness subdimension of timestamps is high. Furthermore, to make time-wise sense from the information that is recorded, timestamps are recorded at the level of milliseconds. So, the granularity subdimension of timestamps is high. However, due to the very precise recording of events, there may be issues with regard to the correctness of the recorded timestamps. For example, for a provider of X-rays all over the world we have seen that the event and its timestamp was recorded at the moment that the storage buffer for logging information was emptied. As a result, the correctness of timestamps is average. All together, the level of abstraction of events is low and the level of accuracy of timestamps can range from average to high.

For solving the type of questions in Section 3, typically data from administrative systems are used. However, for this data, the timestamps of events are only registered in days and the level of abstraction of events is high. This may cause problems when answering the frequently posed questions in Section 3. For example, for discovering the most followed paths as part of question type "Q1", the exact ordering ordering of some activities may not be clear due to the fact that precise timing information is not available. As a result, activities in the discovered process may occur in parallel while in reality this might not be the case. Subsequently, this may also cause issues when comparing processes for different patient groups in order to answer questions of type "Q2". Besides, checking the compliance with internal and external guidelines as part of question type "Q3" might be problematic. Often, detailed knowledge about the process steps that are performed at a medical department is required for compliance checking. As the level of abstraction of data coming from administrative systems is high, this detailed information may not be available and therefore problems during compliance checking are faced. Furthermore, for obtaining performance information about the discovered process as part of question type "Q4" the high level of abstraction of events and the high level of accuracy of timestamps may both

be problematic. For example, highly accurate timestamps on a task level are required for gaining insights into in-hospital waiting times.

So, in order to provide a complete answer to the four types of questions provided in Section 3, data from other systems of the spectrum of Figure 2 is required. For example, data from a healthcare logistics system may be required in order to obtain the exact appointment time of an examination. Also, data from a clinical support system may be taken in order to discover the steps that are typically performed at a medical department as part of an examination.

5 Case Study

In this section, a case study is discussed in order to illustrate the challenges that are faced when answering frequently posed questions during process mining projects. The case study has been performed at the gastro-enterology department of azM and involves a group of colorectal cancer patients for which surgery was needed. First, we discuss the questions that have been asked by the medical professionals. Second, we present the data that has been used for solving these questions. Third, results and open challenges are discussed.

In Section 3, we introduced questions Q1 to Q4. AzM was particularly interested in *Q1: what are the main paths that are followed by patients in the process?*, *Q3: do we comply with the national guideline regarding the treatment of cancer patients?* and *Q4: are there medical disciplines for which long waiting times exist?*. One of the main reasons that these questions were asked was that in January 2009, a dedicated nurse has been appointed which is responsible for the efficient planning of appointments and for acceptable throughput times.

As the above mentioned questions relate to the entire process of diagnosing and treating patients, we collected data from an "administrative system" within azM. This system takes care of the fact that by Dutch law all hospitals need to record the diagnosis and treatment steps at the level of individual patients in order to receive payment. Using this system, we extracted data of gastro-enterology patients which have been treated for rectum cancer from 2009 till 2012 and for which surgery was needed. A snippet of the data that we received is shown in Figure 3. Here, each line describes a service that has been delivered to a patient. Note that the data has been anonymized in order to maintain confidentiality. The first line shows that the mean corpuscular volume of the blood was calculated (column "description operation") which was requested by doctor 'Mans' (column requesting relation) and performed by doctor "Vanwersch" (column "executing doctor") from the haematological lab (column "description department") on July 10th 2008 (column "start date operation"). Note that for each service delivered, it is only known on which *day* the service has been delivered.

Figure 4 shows some results regarding the process till surgery. These results have been used for answering the above mentioned questions. Figure 4a, shows a dotted chart which visualizes events as dots. On the vertical axis the different cases (i.e. patients) are shown and events are colored according to their activity

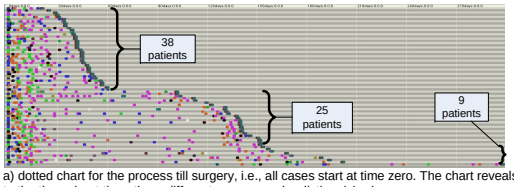
patient identifier	day of birth	sex	address	requesting relation	executing doctor	requesting depart-ment	executing department	description department	operation	description operation

	A	B	C	D	E	F	G	H	I	J	K
1	Patient	Geboorte datum	Ge slacht	BAS: plaats	Aanvragende Relatie	Uitvoerend Arts	Aanvr. verpl. OE	Uitvoerende OE	Beschrijving OE	Ver richtin	Beschrijving verrichting
2	999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	678700	m.c.v. mean corpusculair v
3	'999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	370407D	hemoglobine foto-elektrisch
4	'999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	370712B	leukocyten tellen elektr.hem
5	'999999	7-11-1950	Man	Maastricht	Mans	Vanwersch	PINT	LHMA	LAB - HAEMATOLOG	370715A	trombocyten tellen-electron
6	'999999	7-11-1950	Man	Maastricht	Mans	LKC algemeer	PINT	LCHE	LAB - KLIN.CHEMIE A.	370423	alkalische fosfatasebloed
7	'999999	7-11-1950	Man	Maastricht	Mans	LKC algemeer	PINT	LCHE	LAB - KLIN.CHEMIE A.	370442	natriumbl

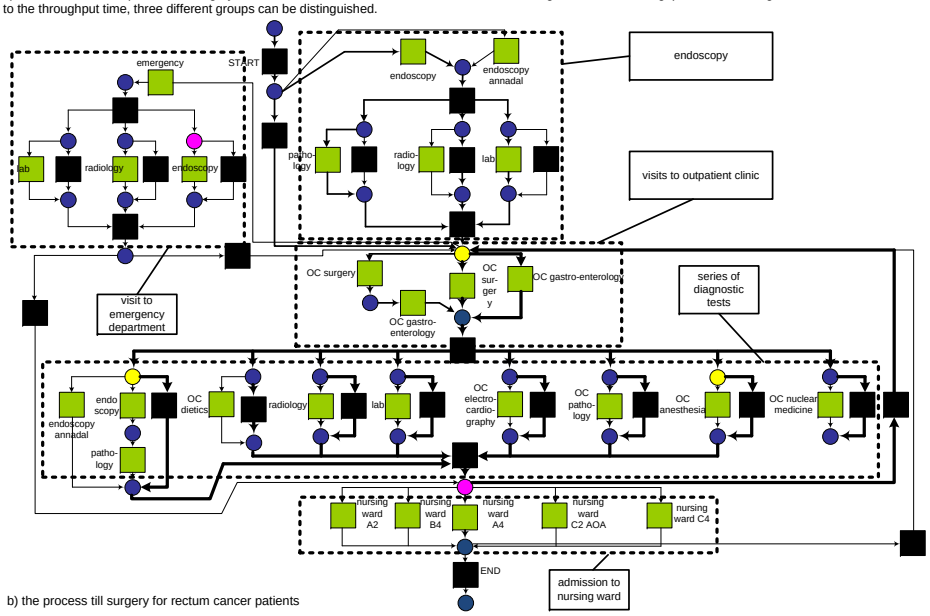
movement identifier	trajectory identifier	external status	trajectory code	start date operation	start year of trajectory	diag-nosis	description diagnosis	start date trajectory	number of operations	department identifier

	L	M	N	O	P	Q	R	S	T	U	V
	Beweging	Traject Nummer	Status extern	Traject Code	Begindatum Verrichtin	Begin jaar	Diagnose	Diagnose omschrijving	Begindatum Traject	Aant. Verricht	OE_ Verrichtin
1	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_678700
2	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_370407D
3	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_370712B
4	1000144779/17	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LHMA_370715A
5	1000144779/14	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LCHE_370423
6	1000144779/14	000001	Gefactureerd	18.11..751..103	13-02-2009	2009	18/751	Acute pancreatitis	10-07-2008	1	LCHE_370442

Fig. 3. A snippet of the excel file that contained the raw data of azM. For each column the Dutch description has been provided.



a) dotted chart for the process till surgery, i.e., all cases start at time zero. The chart reveals a large variation in throughput. Also, with regard to the throughput time, three different groups can be distinguished.



b) the process till surgery for rectum cancer patients

Fig. 4. A dotted chart and Petri net describing the patient process before surgery.

names. As can be seen, the process is shown using relative time, i.e. all cases start at time zero. The chart shows that there is a large variation in the total throughput time of cases. Furthermore, three patient groups can be distinguished. The first group of 38 patients did not receive any radiotherapy before surgery. The second group of 25 patients received radiotherapy before the surgery explaining the longer throughput times (from 101 days till 154 days). The last group of 9 patients appeared to be complex cases for which an individualized treatment was necessary.

For the first group of patients, the total throughput times ranges from 21 days to 60 days. As a consequence, the national guideline regarding acceptable waiting times between diagnosis and treatment for cancer patients is violated. The discovered process is shown in the Petri net of Figure 4b. For the process model the aim was to extend it with performance information. Therefore, it was required that the model is a good reflection of the behavior captured in the log. The following approach has been used for the construction of the model. First, a process mining algorithm has been applied in order to discover a process model which shows the medical departments that were visited and their order. Second, the process model was adapted by hand and it was checked in the process mining tool ProM how well it reflected the behavior in the log. This second step was repeated till the model was a good reflection of the behavior captured in the log. Furthermore, performance information has been projected on the model by coloring the places. A blue color indicates a low waiting time (less than 5 days), a yellow color indicates a medium waiting time (between 5 and 10 days), and a pink color indicates a high waiting time (more than 10 days). The thickness of the arc indicates how often the path has been followed, i.e a thick arc means that the path has been followed often. Note that the grey rectangles represent medical departments whereas the black rectangles are only added because of routing purposes. In general, the process is as follows. First, an endoscopy takes place followed by some diagnostic tests (e.g. a radiology test) or the patient immediately visits the outpatient clinic of gastro-enterology. After visiting the outpatient clinic, a series of diagnostic tests takes place (e.g. a lab or radiology test) followed by another contact of the patient with a doctor of either gastro-enterology or surgery, i.e. a visit to the hospital or a consultation by telephone. Also, after the tests, it may occur that the patient suddenly visits the emergency department of the hospital or the patient may be admitted to the hospital.

Figure 4b shows that there is a high waiting time before the patient can be admitted to the hospital (average: 12.87 days, standard deviation: 5.68 days) and before a patient has contact with a doctor of gastro-enterology or surgery. For the outpatient clinic of gastro-enterology the waiting time is on average 4.25 days (standard deviation: 15.32 hours) whereas the waiting time for surgery is on average 6.63 days (standard deviation: 16.61 hours). Furthermore, the process of having a contact with the doctor followed by a series of diagnostic tests is repeated multiple times. On average, this is repeated more than two times.

In conclusion, clear suggestions can be provided for improving the process. However, there are still some challenges remaining due to limitations of the

provided data. For example, there is quite some parallelism in the process of Figure 4b which causes problems when answering the question about the main paths that are followed by patients in the process (question type "Q1"). As only day timestamps were available for the events, it was difficult to find more causal relations. Regarding compliance checking (question type "Q3") it was not possible to check the compliance with the national guideline for access times due to the fact that information about scheduling of appointments is missing in an administrative system. Also, if a patient has multiple appointments on one day, it cannot be seen how much in-hospital waiting time passes between these appointments. The latter causes problems in answering the question about the medical disciplines for which long in-hospital waiting times exist (question type "Q4"). In order to solve the aforementioned issues it is required to augment the available data with data from other systems (e.g. from a healthcare logistics system).

6 Conclusions

In this paper we have discussed the application of process mining in hospitals. Therefore, we have first given an overview of the questions that are frequently posed by medical professionals. Afterwards, by means of a spectrum, we have described the different types of event data found in current HISs and elaborated upon whether this event data allows for answering the questions posed in process mining projects.

Although process mining has matured over the last years and many mining techniques are available, several important open problems and challenges are remaining. First, when looking to all four systems of the spectrum, it is clear that data is spread around disparate data sources. Based on the questions posed by the medical professionals, data may be required from different data sources. This requires that links between the four systems of the spectrum are clear. The usage of ontologies is interesting in this respect. Ontologies can be used for defining an appropriate scope and to identify the case from the data sources. Currently, some research is performed in the context of ontologies and (semantic) process mining [3, 20]. Further research should explore opportunities for developing an ontology for process mining systems in the healthcare domain.

Second, several problems and challenges can be distinguished when looking to individual systems of the spectrum. For example, for many types of systems we have to deal with events of which the granularity of the timestamp is low. One reason is that the timestamp only refers to the day on which events occur. Consequently, current process mining algorithms have problems with identifying the correct control-flow as the ordering of events within the log do not necessarily conform to the ordering of events on the day itself. Further research is needed in this respect.

Another issue is that a timestamp may not be correct based on the chosen level of granularity. Currently, some research has been performed (also in a healthcare context) on the identification of abnormal cases and infrequent exe-

cution patterns by means of outlier detection techniques [4, 7]. Future research is needed in order to identify suspicious patterns regarding the recording of timestamps of events, not only within the same case but also among multiple cases.

Acknowledgements

This research is supported by the Dutch Technology Foundation STW, applied science division of NWO and the Technology Program of the Ministry of Economic Affairs.

References

1. W.M.P. van der Aalst. *Process Mining: Discovery, Conformance and Enhancement of Business Processes*. Springer-Verlag, Berlin, 2011.
2. T. Blum, N. Padoy, H. Feuner, and N. Navab. Workflow Mining for Visualization and Analysis of Surgeries. *International Journal of Computer Assisted Radiology and Surgery*, 3:379–386, 2008.
3. A.K. Alves de Medeiros and W.M.P. van der Aalst. *Advances in Web Semantics I*, volume 4891 of *Lecture Notes in Computer Science*, chapter Process Mining towards Semantics, pages 35–80. Springer-Verlag, Berlin, 2008.
4. L. Ghionna, G. Greco, A. Guzzo, and L. Pontieri. Outlier Detection Techniques for Process Mining Applications. In *ISMIS 2008*, pages 150–159, 2008.
5. C.W. Günther, A. Rozinat, W.M.P. van der Aalst, and K. van Uden. Monitoring Deployed Application Usage with Process Mining. BPM Center Report BPM-08-11, BPMcenter.org, 2008.
6. S. Gupta. Workflow and Process Mining in Healthcare. Master's thesis, Eindhoven University of Technology, Eindhoven, 2007.
7. B. Han, L. Jiang, and H. Cai. Abnormal Process Instances Identification Method in Healthcare Environment. In *Proceedings of IEEE TrustCom 2011*, pages 1387–1392, 2011.
8. R. Haux, A. Winter, E. Ammenwerth, and B. Brigl. *Strategic Information Management in Hospitals: An Introduction to Hospital Information Systems*. Springer-Verlag, Berlin, 2004.
9. M. Lang, T. Bürkle, S. Laumann, and H.-U. Prokosch. Process Mining for Clinical Workflows: Challenges and Current Limitations. In *Proceedings of MIE 2008*, volume 136 of *Studies in Health Technology and Informatics*, pages 229–234. IOS Press, 2008.
10. R. Lenz and M. Reichert. IT Support for Healthcare Processes - Premises, Challenges, Perspectives. *Data and Knowledge Engineering*, 61:49–58, 2007.
11. R.S. Mans, M.H. Schonenberg, G. Leonardi, S. Panzarasa, S. Quaglini, and W.M.P. van der Aalst. Process Mining Techniques : An Application to Stroke Care. In *Proceedings of MIE 2008*, volume 136 of *Studies in Health Technology and Informatics*, pages 573–578. IOS Press, 2008.
12. R.S. Mans, M.H. Schonenberg, M.S. Song, W.M.P. van der Aalst, and P.J.M. Bakker. Application of Process Mining in Healthcare : a Case Study in a Dutch Hospital. In *Proceedings of BIOSTEC 2008*, volume 25 of *Communications in Computer and Information Science*, pages 425–438. Springer-Verlag, Berlin, 2009.

13. L. Perimal-Lewis, S. Qin, C. Thompson, and P. Hakendorf. Gaining Insight from Patient Journey Data using a Process-Oriented Analysis Approach. In *HIKM 2012*, volume 129 of *Conferences in Research and Practice in Information Technology*, pages 59–66. Australian Computer Society, Inc., 2012.
14. J. Poelmans, G. Dedene, G. Verheyden, H. van der Mussele, S. Viaene, and E. Peters. Combining Business Process and Data Discovery Techniques for Analyzing and Improving Integrated Care Pathways. In *Proceedings of ICDM'10*, volume 6171 of *Lecture Notes in Computer Science*, pages 505–517. Springer-Verlag, Berlin, 2010.
15. R. Rada. *Information Systems and Healthcare Enterprises*. IGI Global, 2008.
16. L. Torres Ramos. Healthcare Process Analysis : Validation and Improvements of a Data-based Method using Process Mining and Visual Analytics. Master's thesis, Eindhoven University of Technology, Eindhoven, 2009.
17. A. Rebuge and D.R. Ferreira. Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining. *Information Systems*, 37(2), 2012.
18. P. Riemers. Process Improvement in Healthcare : a Data-based Method using a Combination of Process Mining and Visual Analytics. Master's thesis, Eindhoven University of Technology, Eindhoven, 2009.
19. Treekoverleg. TR-039, Notitie Streefnormstelling Wachttijden Curatieve Sector, 2000. In Dutch.
20. J.M.E.M. van der Werf, H.M.W. Verbeek, and W.M.P. van der Aalst. Context-Aware Compliance Checking. To appear in Proceedings of BPM 2012.
21. J.Y. Zhou. Process mining : Acquiring Objective Process Information for Healthcare Process Management with the CRISP-DM Framework. Master's thesis, Eindhoven University of Technology, Eindhoven, 2009.