

A guide for the application of analytics on healthcare processes: a dynamic view on patient pathways

Jasmien Lismont^{a,1,*}, Anne-Sophie Janssens^a, Irina Odnoletkova^b, Seppe vanden Broucke^a, Filip Caron^a, Jan Vanthienen^a

^a*KU Leuven, Dept. of Decision Sciences and Information Management, Naamsestraat 69, B-3000 Leuven, Belgium*

^b*KU Leuven, Dept. of Public Health, Kapucijnenvoer 35, B-3000 Leuven, Belgium*

*Corresponding author

Email address: Jasmien.Lismont@kuleuven.be (Jasmien Lismont)

¹Phone: +3216321775

Abstract

Objective: The aim of this study is to guide healthcare instances in applying process analytics on healthcare processes. Process analytics techniques can offer new insights in patient pathways, workflow processes, adherence to medical guidelines and compliance with clinical pathways, but also bring along specific challenges which will be examined and addressed in this paper.

Methods: The following methodology is proposed: log preparation; log inspection; abstraction and selection; clustering; process mining; and validation. It was applied on a case study in the type 2 diabetes mellitus domain.

Results: Several data pre-processing steps are applied and clarify the usefulness of process analytics in a healthcare setting. Healthcare utilization, such as diabetes education, is analyzed and compared with diabetes guidelines. Furthermore, we take a look at the organizational perspective and the central role of the GP. This research addresses four challenges: healthcare processes are often patient and hospital specific which leads to unique traces and unstructured processes; data is not recorded in the right format, with the right level of abstraction and time granularity; an overflow of medical activities may cloud the analysis; and analysts need to deal with data not recorded for this purpose. These challenges complicate the application of process analytics. It is explained how our methodology takes them into account.

Conclusion: Process analytics offers new insights into the medical services patients follow, how medical resources relate to each other and whether patients and healthcare processes comply with guidelines and regulations.

Keywords: Patient pathway, Process analytics, Healthcare processes, Electronic healthcare records, Case study

1. Introduction

Recently, patients and healthcare processes are increasingly more tracked in electronic healthcare records (EHR) creating a vast amount of data [1]. These data can provide us with new insights in applications such as

- discovering the most frequent and exceptional paths of medical services;
- comparing pathways of different patient groups;
- modeling medical resources, e.g. a GP, involved in treatments and analyzing potential bottlenecks in the workflow;
- analyzing whether patients follow healthcare guidelines, e.g. the American Diabetes Association (ADA) guidelines, and whether the process complies with healthcare regulations.

More specifically, process analytics (PA) techniques can help us acquire these insights because of the dynamic perspective they offer on captured reality. This article focuses on the benefits PA can present to the healthcare environment and more specifically the diabetes domain. Type 2 diabetes mellitus (T2DM) is a growing and very present disease in our society [2]. Moreover, it raises additional challenges both from a social and from an informatics viewpoint.

This research presents a novel guide for the application of PA in healthcare. It was applied on a case study in the diabetes care sector and identifies the opportunities and challenges involved. Section 2 provides an introduction to PA. Consecutively, the case study is described in section 3 and the specific challenges of applying PA in healthcare are discussed in section 4. The methods are explained in section 5 and applied to the case study in section 6. Finally, the results are discussed in section 7.

2. Process analytics

Process analytics. This domain refers to the advanced analysis of time-oriented data to derive new insights. In healthcare, one can define this process of a patient as the different steps, treatments and drugs he takes, which is sometimes also referred to as the patient's 'care journey'. In general, there are three main streams of application of PA: process discovery, conformance analysis, and process extension [3]. The former relates to descriptive techniques to evaluate patients' healthcare steps, the timing and duration of medical services and possibly the medical resources involved. Conformance

analysis can involve adherence to guidelines and regulations. Extension, on the other hand, enhances an existing process model, e.g. a clinical pathway [4]. For each stream of applications, one can focus on a control-flow, organizational or case (e.g. the patient) perspective. For more information on PA, the reader is referred to van der Aalst [3] and Mans et al. [5]. The added value of process analytics in healthcare was already illustrated by Yoo et al. [6] who specifically suggest its suitability for care processes and performance analysis. Other healthcare applications can be found in Mans et al. [7] and Mans et al. [5].

Tools. Two PA tools, Disco and ProM, were applied in this research. Disco [8] is a commercial tool focusing on visualization and user-friendliness. ProM [9] is an open source tool that works with plug-ins which implement PA techniques.

3. Case description

Our study is carried out on a dataset provided by the Independent Sickness Funds of Belgium (MLOZ). In Belgium, health insurance is obligated for all its residents and is financed by means of social contributions of employers and employees. All contributions and reimbursements are controlled by the Rijksinstituut voor Ziekte- en Invaliditeitsverzekering (RIZIV) through several sickness funds. MLOZ followed patients with T2DM in the context of a nurse-led telecoaching project [10] and provided us with the medical services these patients received.

The data was received in the form of several csv files which were combined into one dataset and prepared as an event log which captures all the activities in which patients are involved. Each row in this file represents an event which is an occurrence of an activity, namely a medical service. This event is characterized by a unique case number (the patient), an activity identifier, a timestamp and optional additional case or activity features. For example, an event can be patient 1 (case) consulting his GP (activity) on the 6th of June, 2011 (timestamp). We also retrieved various patient data such as age and gender, as well as data on resource utilization and the costs of care. In total there are 571 patients. Patients range between 18 and 75 years old and were diagnosed with T2DM. They were followed during one year, starting in 2011. The last registered event occurred in 2013. The granularity of the timestamp is on the level of a day. This means that we cannot distinguish a sequence of events in the same day. Arbitrarily, we assume the order of entry as the order of the events per day. Figure 1 provides an extract of the event log as an illustration. The activities in this context are all medical

services the patient received including but not limited to diabetic related services. Medical services are identified by a 6-digit standardized nomenclature code defined by the RIZIV² which captures the events only for the purpose of the Belgian national reimbursement scheme of healthcare insurance, e.g. code 102874 stands for an endocrinologist consultation. In this sense event codes can on the one hand be very detailed (if different costs are associated) and on the other hand too generic (e.g. not specifying the subject of a GP visit). All the services are subsequently divided into categories, identified by Odnoletkova [11]. These categories can be general healthcare services or diabetes specific healthcare services and are based on a protocol [11] and a report on diabetes by the ‘Intermutualistisch Agentschap’. In an event log, one can follow a patient over time by filtering the events per case. This path of activities a patient follows, a patient’s care journey, is referred to as a trace, see Figure 1.

4. Challenges

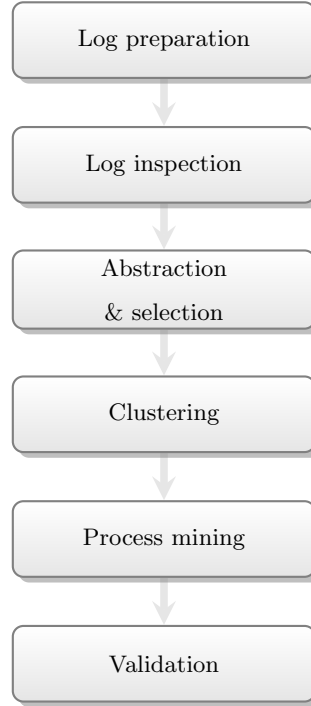
Rebuge and Ferreira [12] identified four challenges in healthcare PA: (1) highly dynamic processes; (2) highly complex processes; (3) increasingly multi-disciplinary processes; and (4) ad-hoc processes.

Additionally, we encountered the following difficulties. Firstly, many researchers assume regular behavior, complemented by process variants, but there is no such thing in our case study. Every patient follows a unique path and is thus accompanied by a unique flow. This is common for chronic diseases, such as diabetes, as patients undergo a variety of medical services during their lifetime. In order to obtain comprehensible models, additional pre-processing of the data is required. Hripcsak et al. [13] stated with regards to this topic that patients in EHRs tend to be heterogeneous, which complicates analyses and increases variance. Secondly, there is an abundance of events since each event is characterized by a numerical RIZIV code. This is a consequence of the fact that we work with data from an administrative system with a low abstraction level [7]. Thirdly, although these nomenclature codes are very detailed in one sense, they are not in another. Most of the codes are not linked to a disease unless the event itself is linked to it. E.g. a visit to the ophthalmologist is not necessarily linked to diabetes. Fourthly, each patient was only followed during one year, which means that we cannot research the entire control flow of one patient.

²See Nomensoft database: <https://www.riziv.fgov.be/webprd/appl/pnomen/Search.aspx?lg=N>

5. Methods

PA can deliver valuable insights into patients' care journeys. However, besides existing challenges, the healthcare domain brings along its own difficulties. Therefore, this research presents a methodology consisting of six steps towards applying PA on healthcare data. This methodology aims to guide users through the various challenges encountered while applying PA to EHRs. Moreover, the different steps must be considered as an iterative process of actions towards well supported conclusions. Klimov et al. [14], as such, explicitly mention the lack of iterative application as a shortcoming of previous studies.



Log preparation. This phase involves the aggregation of event data from different data sources [12] into one event log. Our case study starts with a prepared event log. The aggregation of EHRs brings along its own challenges [15, 16, 17]. Nevertheless, PA techniques can help identifying and solving data quality issues.

Log inspection. This phase relates to the first investigation of the data by means of statistics [12] and simple PA techniques. The goal of this step is to fully understand the dataset and its limitations and opportunities. Large abnormalities can already be discovered.

As such, one can analyze which medical resources provided the most medical services and how many medical services the average patient received. This already provides an abundance of information for health insurances and medical instances. A common concept in PA is the process map which visualizes the patient’s care journey. This could be linked to case frequency, i.e. how often a patient received a service at least once, and absolute frequency, i.e. how often a service is received in absolute terms.

Abstraction and selection. The abstraction and selection phase is a necessary step for unstructured processes.

Abstraction omits irrelevant details and helps to focus on the most significant aspects of the process. One solution is to include expert knowledge. Based on a protocol [11] and on communication with a medical expert, we extracted knowledge about the categories of interest and their nomenclature codes. Hereafter, we replaced the activities with this abstraction. Remember, however, that one can no longer discover paths between activities within these categories. Another solution abstracts activities automatically. Preferably, expert knowledge is still needed to select the patterns of interest. The pattern abstraction technique [18] groups activities that frequently appear together. The user can decide to abstract these patterns into a new activity in the event log. This technique can also be applied iteratively, when the new event log is used as input for the plug-in. Another technique is activity clustering mining [19]. This technique identifies activities that always happen in series and that share some characteristics such as identical medical service providers. Both automated abstraction techniques were selected because of their applicability on healthcare datasets, as explained in Table 1, and because of their focus on the dynamical aspect of the dataset.

Selection techniques are interesting to focus on a subpart of the event log which often leads to more structured results. Event logs were filtered per subgroup for certain categories, such as education in order to compare patients that follow certain guidelines with patients who do not, or such as ‘ophthalmology’ to focus on this aspect of the diabetes process.

Clustering. This phase divides patient traces into separate segments with their own characteristics. A sub-log is created by clustering the traces in groups that have very similar characteristics. Patients are thus clustered not (only) based upon the patient data but on the characteristics of their care journeys. There are two important clustering methods from a process perspective [12], namely trace clustering and sequence clustering. Trace clustering clusters care journeys based upon trace profiles [20]. A trace profile contains certain characteristics of the trace that one finds important, e.g. activities,

a certain sequence of activities, medical service providers, etc. ActiTrac [21] has implemented this technique. ActiTrac is particularly useful because it explicitly considers fitness measures, because of its visualization properties, and because of its promising results in a previous benchmark study [22]. Sequence clustering focuses only on the sequence of activities in traces and creates more simple models than trace clustering [12]. Clustering not only allows to discover more comprehensible models but also allows to identify or confirm subgroups based upon their behavior.

Process mining. Once the event log is abstracted and clustered, the actual process mining can start. We refer to van der Aalst [3] for a detailed overview of process mining techniques. The techniques presented in this paper are the dotted chart analysis [23] and the fuzzy miner [24] from a process perspective, and the role hierarchy plug-in [[25] in [26]] and the social network miner [27] from an organizational viewpoint. The dotted chart can provide a quick visualization of the spread of (a category) of events. This technique was explicitly selected for its visual and comprehensible outcome which, furthermore, enhances communication with healthcare experts. This may be used for checking compliance to medical guidelines or for optimization of resources and processes. The fuzzy miner is particularly useful, especially in a healthcare context, because it is able to cluster events in the process map, thus providing a more comprehensive process model on a higher abstraction level. It answers questions such as: ‘Which services are related (clusters)?’, ‘Which services are followed or preceded by which other services (paths)?’ and ‘How frequently are services consumed?’ Moreover, in an earlier benchmark study [22], the cluster clarity and cluster detection capabilities of this techniques were already illustrated. The role hierarchy plug-in is very useful to observe which resources provide a service the most and the technique is convenient for compliance to regulations. The social network miner plug-in is interesting for analyzing the dynamics between the medical resources and to improve the management of internal relationships.

Validation. Finally the results must be presented to the healthcare instance and validated by experts. Moreover, close contact with experts during the execution of the previous steps is meaningful in order to guide the analysis.

Techniques for each step of the methodology were taken from the literature and can be consulted in Table 1. Except when explicitly mentioned, they are all implemented in the ProM toolkit. The categorization of techniques is not entirely disjunct as some techniques can be applied with different

goals in mind.

6. Results

6.1. Log inspection

We can visualize the activities by means of a process map. In Figure 2, for example, one can observe a high use of ambulatory laboratory tests and ambulatory medication. Physiotherapy (22%), orthopedic services (10%) and ophthalmologic services (1%) are not consulted by many patients during that year, although a yearly visit to an ophthalmologist is recommended by ADA for patients suffering from T2DM. This process map was discovered on an event log with activities abstracted in general healthcare activities [11], see Section 5 ‘Abstraction and selection’, and as such illustrates the importance of an iterative approach. The arcs in the process map represent the most commonly followed paths and are weighted with the number of patients that have at least once taken the path. The nodes represent the activities weighted by the number of patients that have at least once received a service from that category. In general, the darker, the more common a node or arc is in a patient’s care journey. The triangle symbol indicates the start of the process map and the rectangle signifies the end.

Furthermore, we can visualize which drugs are frequently consumed and if they are often taken in combination with other drugs. In Figure 3 one can observe that blood glucose lowering drugs (A10B) are regularly succeeded by or taken concurrently with beta blocking agents (C07), ACE-inhibitors (C09) or lipid modifying agents (C10). This process map was discovered on an event log filtered on medication events, see Section 5 ‘Abstraction and selection’.

6.2. Abstraction and selection

We applied the pattern abstraction [18] plug-in on diabetes specific activities. This suggested, for example, the pattern ‘Cholesterol in serum’, ‘Creatinine in serum’ and ‘HbA1c’. Furthermore, the activity clustering miner [19] was applied. This provided minimal conflict-free sets³. As such, one cluster aggregates the following activities: ‘fixed-rate honorarium radiology’; ‘mammography per breast and armpit’; ‘ultrasound of one or both breast’; and ‘consult ambulant radiology’.

³Set is used to denote a group of clusters of events. Clusters are said to be in conflict if they share at least one identical event. For more information the reader is referred to Günther and van der Aalst [19].

6.3. Clustering

We applied ActiTrac [21] with the maximal amount of clusters set to 10. The result was 10 clusters of which 9 contain one trace and the rest category contains the remaining 562 traces, an indication that the degree of trace variants is extremely high. As such, ActiTrac aims at clustering the typical behavior in the first 9 traces and the remainder in a final rest cluster. However, in this application no typical behavior could be found, i.e. each trace is unique. This indicates that the discovered process model is not well able to capture the behavior contained in the event log. We repeated this analysis on the event log with a higher abstraction level which provided similar results.

6.4. Process mining

Diabetes education analysis. Given the complexity of T2DM, it is recommended by ADA that patients follow diabetes education in order to prevent further progression. We applied a dotted chart analysis [23] on education events of a specific group of patients. The dotted chart is represented in Figure 4. In the dotted chart one can observe the patients on the vertical axis and the days on the horizontal axis. The dots represent consumptions of education services. The following activities are relevant (based on [11]): ‘Follow-up of the patient on insulin’; ‘Education of diabetes patients’; ‘Education in case of complications’; ‘Yearly rate for education of care trajectory patients’; ‘Increased yearly rate for patients who start self-regulation’ and ‘Yearly rate for general coaching’. The three latter activities are billing activities. A relative time line was used, so every first service starts at time zero.

One notices that more than half of the patients follow diabetes education activities only once or twice a year. Moreover, these events are often billing services. There are some patients who follow diabetes education more frequently, but with not much days in between the activities. As such, the average spread between education services per patient in this group is 11.18 days. These activities are mostly education of diabetes patients.

Organizational perspective. The role hierarchy miner plug-in [[25] in [26]] illustrates which provider delivers which services. As the system registers only legal provider-service combinations, no surprising combinations are found. However, not all combinations make sense at first glance. For example, the combination of provider ‘anesthetist and cardiologist’ with the activity ‘Orthopedic shoes and artificial limbs’ activity seems curious. Drilling down on this service gives two times an event which stands for special running shoes, an activity that should be under supervision of an orthopedist. As such, this techniques proves its usefulness for conformance checking to hospital or regulatory policies.

The handover of work between providers can be analyzed by means of the social network miner [27]. Figure 5 (a) shows the handover of work map which only includes connections with a weight of at least 10%. This weight represents the ratio of the total amount of times a transfer occurred and the number of medical service providers that could have succeeded the provider of the transfer. One can observe that the GP is a center point who refers to a lot of specialists and that patients return to their GP after their visit to a specialist. Specialists are also handing over work to themselves or similar specialists, as shown in Figure 5 (b).

Analysis of the event log by means of fuzzy miner. Fuzzy miner [24] was applied on a log which contains no events related to drug intake and which was pre-processed by means of sequence clustering, see Section 5 ‘Clustering’. The resulting fuzzy maps illustrate that the most important activities are on one side basic healthcare activities like wound care or first aid and on the other side dosing and counting activities, which are related to each other through various types of tests. For an illustration of a fuzzy map we refer to Figure 6.

7. Discussion

The *log inspection* phase provided interesting visualization results. The process map already indicates that the diabetes guidelines developed by ADA were not well followed regarding ophthalmology. Another process map indicated that almost all patients took at least once blood glucose lowering drugs, which is sensible for diabetic patients, and that they often take consecutively or concurrently beta blocking agents, ACE-inhibitors or lipid modifying agents, which are often taken in combination (e.g. C10B). After the log inspection phase, the *abstraction and selection* phase follows. These techniques mainly focus on pre-processing, but can already provide meaningful results. For example, activity clustering mining aggregates services that happen closely together, such as a mammography. Hence, one can investigate which services are performed on the same day and by the same department in order to optimize workflow processes, an important part of modern health information systems [28]. The *clustering* phase aims to pre-process the event log, but can also deliver promising results by suggesting subgroups within patient behavior. For example, one could check if the patients of the different clusters share some characteristics such as the stage of their disease. The *process mining* step suggested interesting results for diabetes education services. A possible explanation is that patients are mostly interested in an intense course of diabetes education in a short period of time. The social

network miner takes an organizational perspective and visualized some medical resources who referred to themselves. This could indicate that the specialists are following up on the patient’s condition. Moreover, the GP seems to still be a center point in the care journey of the patients.

7.1. Contribution of the research: addressed challenges

The contribution of this research is the proposal of a guide designed to deal with the specific challenges of PA in healthcare. The following challenges were met during the case study and it is discussed how our methodology can contribute to each of them.

Addressed challenge 1: irregular behavior. There is not always a logical sequence in the services patients consume due to patient-specific treatments. Especially for chronic diseases the paths of services become very irregular and unique. This irregularity can be quickly spotted in the log inspection phase. Once the problem is determined, an abstraction and selection step could be applied. The pattern abstraction plug-in [18] or activity clustering miner plug-in [19] can be applied to abstract services and their internal paths into higher level categories. Selection can be applied to focus on a subset of the behavior, e.g. diabetes education behavior. An alternative or complementary step is to opt for a clustering phase in order to cluster similar traces. If the traces are not similar enough, not much gain will be achieved if one does not carefully apply abstraction and selection first. Moreover, it is advised to focus on clustering traces based on other characteristics than the sequence of services as can be done by means of trace clustering. Often patients do follow similar events, but in other sequences which makes their behavior so unique.

Addressed challenge 2: data abstraction levels and time granularity. Data that is not collected for the purpose of analytics can raise additional challenges. We experienced highly detailed data with a low time granularity. If data can be prepared or registered in an adjusted way, this problem can potentially be solved in the log preparation phase. Too detailed events can be abstracted or a filtering technique can be applied. It is recommended to combine abstraction and selection with process mining plug-ins such as fuzzy miner to improve results even further. Unfortunately, there has not been much research done on low time granularity. Lakshmanan et al. [29] propose a solution which could potentially be further generalized. Additionally, events can be modeled as a “series of sequential activities” [30] when the ordering of these events is not of importance.

Addressed challenge 3: an abundance of events. An abundance of events is often due to irregular behavior and/or too low abstraction levels of the activities. If the problem continues after abstraction and clustering techniques, further selection steps are recommended. Experts can consecutively apply our methodology for each subgroup of either events or patients whereupon results can be aggregated.

Addressed challenge 4: data quality. Hripcsak and Albers [16] summarize four common problems with EHRs: completeness, accuracy, complexity and bias and they, moreover, express a need for a better understanding of these challenges. Also in our dataset we encountered these problems due to the fact that the data is aggregated from multiple health institutions, the high complexity of the processes, the high use of narrative notes and text data and, most importantly, the fact that the data was collected for the purpose of reimbursement. Specifically in the T2DM domain there is some disagreement on treatment practices, guidelines and comorbidities of the disease which leads to an even higher complexity. For this purpose, the methodology can be applied in a different manner. Instead of applying techniques on data to achieve new insights in patient processes, new insights in the data and data quality can be obtained. The abstraction plug-ins can suggest similar services or common paths of treatments to medical researchers. Clustering can offer additional insights to experts regarding comorbidities indicating which treatments are frequently offered to patients. Finally, the process mining step itself can offer new insights and suggest data quality problems. In addition, PA techniques can help to develop and validate clinical pathways.

7.2. Recommendations for future work

First, some of the PA techniques could be further adjusted to healthcare processes. For example, natural language processing techniques could improve abstraction techniques where event logs contain highly textual data. Secondly, we remark some interesting features for enhancement of the process maps. As such, it can be interesting to add a cost perspective of activities alongside the case and absolute frequencies. Potentially this could be complemented by other customizable perspectives such as a focus on drug consummation. Finally, literature suggests some data mining techniques to analyze certain aspects of healthcare. There are techniques that were composed to deal with unstructured and low abstraction level data [31]. Nevertheless, these techniques are not able to deal with the dynamic nature of processes in the way PA does. A combination of these two fields could improve the results. We believe a further elaboration of the use of case data in order to better predict medical services or even costs would improve the results. The inclusion of the cost perspective in PA techniques, where

one links costs to medical services and tries to predict medical services and their costs based on the previous path of a patient (in addition with case data), could deliver promising results to the healthcare and diabetes domain of today.

8. Conclusion

PA is a relatively new research field with already promising results for the healthcare environment [32, 31, 33, 29, 34, 7, 12, 6]. By obtaining deeper insights into medical processes, medical instances such as hospitals and health insurance companies can focus on the optimization of processes, predicting patient outcomes and costs, resource optimization and compliance to patient guidelines or medical regulations. We highly encourage further research and application of PA in healthcare to explore these opportunities. In the view of the aging population, an increasing demand of healthcare services and a growing availability of data, PA has a good potential to support healthcare optimization.

Ethical issues

All patients signed a consent letter with MLOZ to allow for the use of their anonymized personal data for research. KU Leuven has a confidentiality agreement with the MLOZ which permits the authors to use the data for the purpose of this research and to publish the results.

Authors' contributions

JL and AJ were responsible for the design and application of the methods as well as the drafting of the paper. IO contributed to the discussion of the data and the added value of the presented techniques to healthcare. SVB, FC and JV contributed to the discussion of the presented process analytics techniques and how the methods address the various challenges mentioned. IO, SVB, FC and JV also reviewed and revised the manuscript. IO contributed as such mainly from a healthcare perspective and SVB, FC and JV from a process analytics perspective.

Acknowledgments

The authors would like to thank the management of the MLOZ for their consent in using the data independently from the company for the purpose of this research.

Funding sources

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

References

- [1] T. B. Murdoch, A. S. Detsky, The inevitable application of big data to health care, *J Am Med Inform Assoc* 309 (13) (2013) 1351–1352. doi:10.1001/jama.2013.393.
- [2] O. Pinhas-Hamiel, P. Zeitler, The global spread of type 2 diabetes mellitus in children and adolescents, *J Pediatr* 146 (5) (2005) 693–700. doi:10.1016/j.jpeds.2004.12.042.
- [3] W. van der Aalst, *Process mining: discovery, conformance and enhancement of business processes*, Springer-Verlag Berlin Heidelberg, Schleiden, Germany, 2011. doi:10.1007/978-3-642-19345-3.
- [4] R. Blaser, M. Schnabel, C. Biber, M. Bäumlein, O. Heger, M. Beyer, E. Opitz, R. Lenz, K. A. Kuhn, Improving pathway compliance and clinician performance by using information technology, *Int J Med Inform* 76 (2) (2007) 151–156. doi:10.1016/j.ijmedinf.2006.07.006.
- [5] R. S. Mans, W. van der Aalst, R. J. Vanwersch, *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*, Springer International Publishing, Cham, Switzerland, 2015. doi:10.1007/978-3-319-16071-9.
- [6] S. Yoo, M. Cho, E. Kim, S. Kim, Y. Sim, D. Yoo, H. Hwang, M. Song, Assessment of hospital processes using a process mining technique: Outpatient process analysis at a tertiary hospital, *Int J Med Inform* 88 (2015) 34–43. doi:10.1016/j.ijmedinf.2015.12.018.
- [7] R. Mans, W. M. P. van der Aalst, R. J. B. Vanwersch, A. J. Moleman, Process mining in healthcare: Data challenges when answering frequently posed questions, in: *Process Support and Knowledge Representation in Health Care - BPM 2012 Joint Workshop, ProHealth 2012/KR4HC 2012*, Tallinn, Estonia, September 3, 2012, Revised Selected Papers, 2012, pp. 140–153. doi:10.1007/978-3-642-36438-9_10.
- [8] C. W. Günther, A. Rozinat, Disco: Discover your processes, *BPM (Demos)* 940 (2012) 40–44.
- [9] B. F. van Dongen, A. K. A. de Medeiros, H. Verbeek, A. Weijters, W. M. van der Aalst, The ProM framework: A new era in process mining tool support, in: *Applications and Theory of Petri Nets 2005*, Springer, 2005, pp. 444–454. doi:10.1007/11494744_25.

- [10] I. Odnoletkova, G. Goderis, F. Nobels, B. Aertgeerts, L. Annemans, D. Ramaekers, Nurse-led telecoaching of people with type 2 diabetes in primary care: rationale, design and baseline data of a randomized controlled trial, *BMC Fam Pract* 15 (24). doi:10.1186/1471-2296-15-24.
- [11] I. Odnoletkova, Telecoaching in diabetes type 2, Tech. rep., MLOZ, <http://www.mloz.be/nl/publicatie/telecoaching-diabetes-type-2> (2014).
- [12] Á. Rebuge, D. R. Ferreira, Business process analysis in healthcare environments: A methodology based on process mining, *Inf Syst* 37 (2) (2012) 99–116. doi:10.1016/j.is.2011.01.003.
- [13] G. Hripcsak, D. J. Albers, A. J. Perotte, Exploiting time in electronic health record correlations, *J Am Med Inform Assoc* 18 (Supplement) (2011) 109–115. doi:10.1136/amiajnl-2011-000463.
- [14] D. Klimov, A. Shknevsky, Y. Shahar, Exploration of patterns predicting renal damage in patients with diabetes type II using a visual temporal analysis laboratory, *J Am Med Inform Assoc* 22 (2) (2015) 275–289. doi:10.1136/amiajnl-2014-002927.
- [15] K. Bernstein, M. Bruun-Rasmussen, S. Vingtoft, S. K. Andersen, C. Nøhr, Modelling and implementing electronic health records in Denmark, *Int J Med Inform* 74 (2) (2005) 213–220. doi:10.1016/j.ijmedinf.2004.07.007.
- [16] G. Hripcsak, D. J. Albers, Next-generation phenotyping of electronic health records, *J Am Med Inform Assoc* 20 (1) (2013) 117–121. doi:10.1136/amiajnl-2012-001145.
- [17] J. Taggart, S.-T. Liaw, H. Yu, Structured data quality reports to improve EHR data quality, *Int J Med Inform* 84 (12) (2015) 1094–1098. doi:10.1016/j.ijmedinf.2015.09.008.
- [18] R. P. J. C. Bose, W. M. P. van der Aalst, Abstractions in process mining: A taxonomy of patterns, in: *Business Process Management, 7th International Conference, BPM 2009, Ulm, Germany, September 8-10, 2009. Proceedings, 2009*, pp. 159–175. doi:10.1007/978-3-642-03848-8_12.
- [19] C. W. Günther, W. M. van der Aalst, Mining activity clusters from low-level event logs, Working Paper WP 165, Beta, Research School for Operations Management and Logistics (2006).
- [20] M. Song, C. W. Günther, W. M. P. van der Aalst, Trace clustering in process mining, in: *Business Process Management Workshops, BPM 2008 International Workshops, Milano, Italy, September 1-4, 2008. Revised Papers, 2008*, pp. 109–120. doi:10.1007/978-3-642-00328-8_11.
- [21] J. De Weerdt, S. K. L. M. vanden Broucke, J. Vanthienen, B. Baesens, Leveraging process discovery with trace clustering and text mining for intelligent analysis of incident management processes, in: *Proceedings*

- of the IEEE Congress on Evolutionary Computation, CEC 2012, Brisbane, Australia, June 10-15, 2012, 2012, pp. 1–8. doi:10.1109/CEC.2012.6256459.
- [22] T. Thaler, S. F. Ternis, P. Fettke, P. Loos, A comparative analysis of process instance cluster techniques, in: Smart Enterprise Engineering: 12. Internationale Tagung Wirtschaftsinformatik, WI 2015, Osnabrück, Germany, March 4-6, 2015, pp. 423–437, <http://aisel.aisnet.org/wi2015/29>.
 - [23] Dotted char analysis, Tech. rep., Eindhoven: Process Mining Group, <http://www.processmining.org/online/dottedchartanalysis> (2009).
 - [24] C. W. Günther, W. M. P. van der Aalst, Fuzzy mining - adaptive process simplification based on multi-perspective metrics, in: Business Process Management, 5th International Conference, BPM 2007, Brisbane, Australia, September 24-28, 2007, Proceedings, 2007, pp. 328–343. doi:10.1007/978-3-540-75183-0_24.
 - [25] M. Bozkaya, J. Gabriels, J. M. E. M. Werf, Process diagnostics: a method based on process mining, in: Information, Process, and Knowledge Management, 2009. eKNOW'09. International Conference on, IEEE, 2009, pp. 22–27. doi:10.1109/eKNOW.2009.29.
 - [26] C. S. Alves, Social network analysis for business process discovery, Ph.D. thesis, Instituto Superior Tecnico, Av. Rovisco Pais, 1 (2010).
 - [27] W. M. P. van der Aalst, H. A. Reijers, M. Song, Discovering social networks from event logs, Comput Support Coop Work 14 (6) (2005) 549–593. doi:10.1007/s10606-005-9005-9.
 - [28] P. Ciccarese, E. Caffi, S. Quaglini, M. Stefanelli, Architectures and tools for innovative health information systems: the guide project, Int J Med Inform 74 (7) (2005) 553–562. doi:10.1016/j.ijmedinf.2005.02.001.
 - [29] G. T. Lakshmanan, S. Rozsnyai, F. Wang, Investigating clinical care pathways correlated with outcomes, in: Business Process Management - 11th International Conference, BPM 2013, Beijing, China, August 26-30, 2013. Proceedings, 2013, pp. 323–338. doi:10.1007/978-3-642-40176-3_27.
 - [30] S. Suriadi, R. Mans, M. T. Wynn, A. Partington, J. Karnon, Measuring patient flow variations: A cross-organisational process mining approach, in: Asia Pacific Business Process Management - Second Asia Pacific Conference, AP-BPM 2014, Brisbane, QLD, Australia, July 3-4, 2014. Proceedings, 2014, pp. 43–58. doi:10.1007/978-3-319-08222-6_4.

- [31] J. De Weerd, F. Caron, J. Vanthienen, B. Baesens, Getting a grasp on clinical pathway data: An approach based on process mining, in: Emerging Trends in Knowledge Discovery and Data Mining - PAKDD 2012 International Workshops: DMHM, GeoDoc, 3Clust, and DSDM, Kuala Lumpur, Malaysia, May 29 - June 1, 2012, Revised Selected Papers, 2012, pp. 22–35. doi:10.1007/978-3-642-36778-6_3.
- [32] F. Caron, J. Vanthienen, K. Vanhaecht, E. van Limbergen, J. De Weerd, B. Baesens, Monitoring care processes in the gynecologic oncology department, *Comput Biol Med* 44 (2014) 88–96. doi:10.1016/j.compbimed.2013.10.015.
- [33] Z. Huang, X. Lu, H. Duan, On mining clinical pathway patterns from medical behaviors, *Artif Intell Med* 56 (1) (2012) 35–50. doi:10.1016/j.artmed.2012.06.002.
- [34] R. S. Mans, H. Schonenberg, M. Song, W. M. P. van der Aalst, P. J. M. Bakker, Application of process mining in healthcare - A case study in a Dutch hospital, in: Biomedical Engineering Systems and Technologies, International Joint Conference, BIOSTEC 2008, Funchal, Madeira, Portugal, January 28-31, 2008, Revised Selected Papers, 2008, pp. 425–438. doi:10.1007/978-3-540-92219-3_32.
- [35] R. Moskovitch, Y. Shahar, Medical temporal-knowledge discovery via temporal abstraction, *AMIA Annual Symposium proceedings / AMIA Symposium* (2009) 452–456 <http://europepmc.org/articles/PMC2815492>.
- [36] G. M. Veiga, D. R. Ferreira, Understanding spaghetti models with sequence clustering for ProM, in: Business Process Management Workshops, BPM 2009 International Workshops, Ulm, Germany, September 7, 2009. Revised Papers, 2009, pp. 92–103. doi:10.1007/978-3-642-12186-9_10.
- [37] F. M. Maggi, M. Dumas, L. García-Bañuelos, M. Montali, Discovering data-aware declarative process models from event logs, in: Business Process Management - 11th International Conference, BPM 2013, Beijing, China, August 26-30, 2013. Proceedings, 2013, pp. 81–96. doi:10.1007/978-3-642-40176-3_8.
- [38] A. Rozinat, W. M. P. van der Aalst, Decision mining in ProM, in: Business Process Management, 4th International Conference, BPM 2006, Vienna, Austria, September 5-7, 2006, Proceedings, 2006, pp. 420–425. doi:10.1007/11841760_33.

Figures

	Studycode	Start date	Activity	Case feature	Activity feature	Patient cost	Medical Performer nr.	Qualification performer	Timestamp
			Code Nomenclature	Group	RIZIV cost				Date service
Case	390	20110606	541612	1	0.99	0	68110925000	993	20110607
	390	20110606	541892	1	0.63	0	68110925000	993	20110607
Event	390	20110606	541052	1	0.79	0	68110925000	993	20110607
Trace	390	20110606	540551	1	0.80	0	68110925000	993	20120426
	2269	20120301	469755	2	61.24	0	10532121000	140	20130226
	2269	20120301	101076	3	21.82	1.50	13499727000	4	20120524
	2269	20120301	101076	3	22.32	1	13668486000	4	20120604
	2269	20120301	101076	3	22.32	1	13668486000	4	20120613

Figure 1: Extract from the event log. (This is a 1.5-column fitting image)

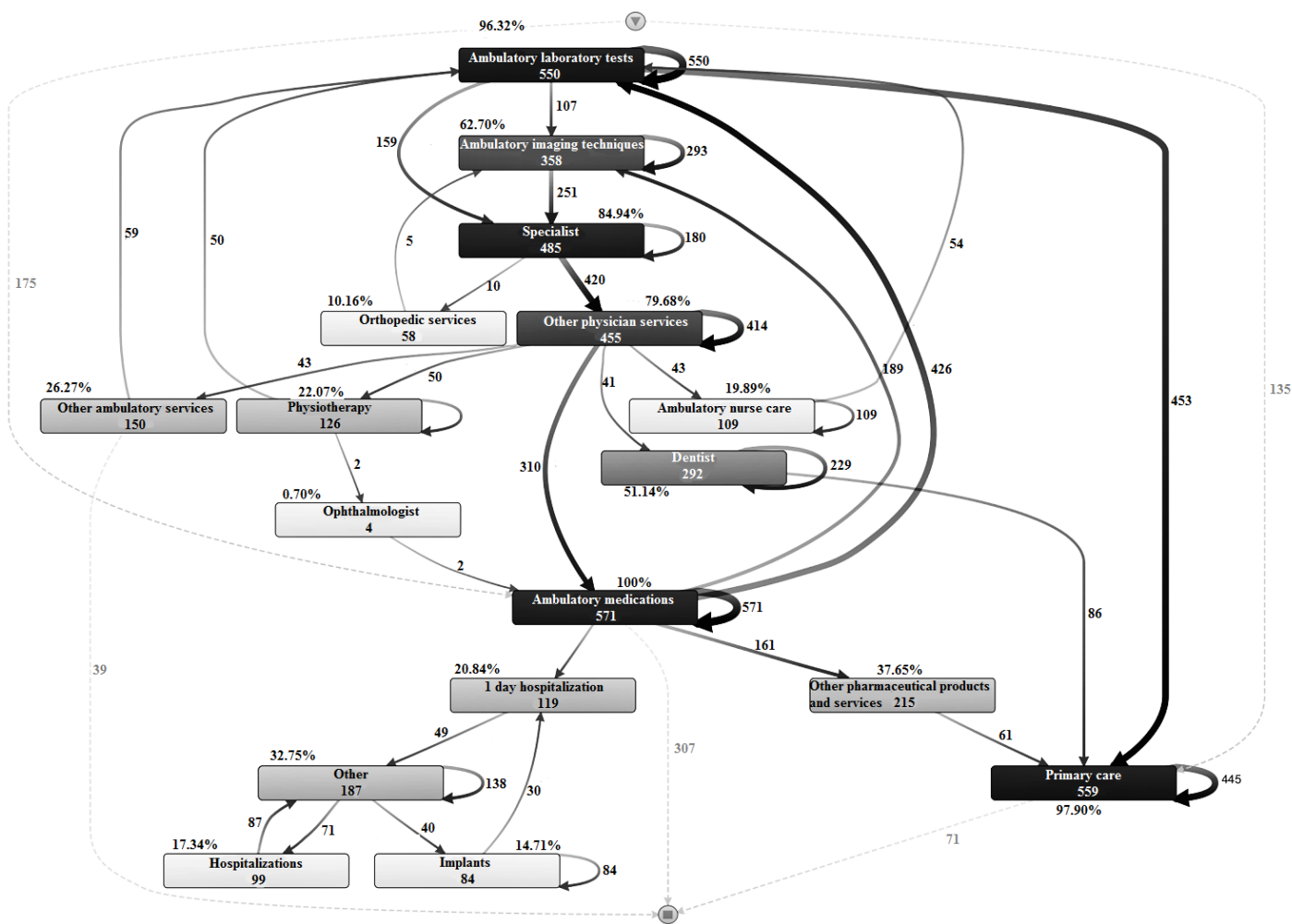
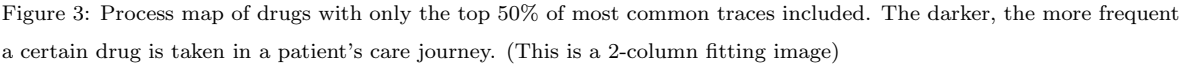


Figure 2: Process map of general healthcare activities with case frequencies (in absolute terms and in percentages) displayed. Only the top 20% most common traces are included. (This is a 2-column fitting image)



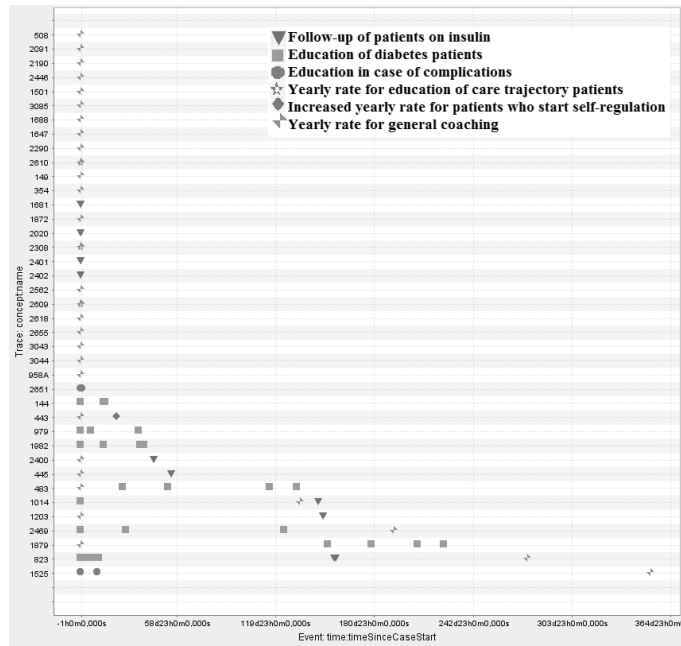


Figure 4: Dotted chart of diabetes education activities. One can observe the patients on the vertical axis and the days on the horizontal axis. The dots represent consumptions of education services. A relative time line was used, so every first service starts at time zero. (This is a single column fitting image)

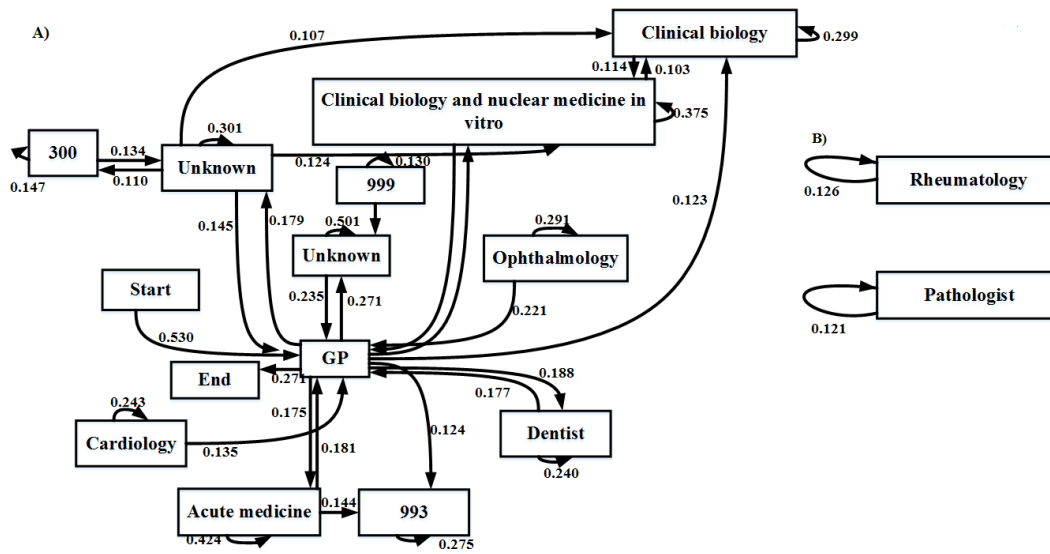


Figure 5: Handover of work: (a) GP performs a central role; (b) follow-up by specialists. (This is a 1.5-column fitting image)

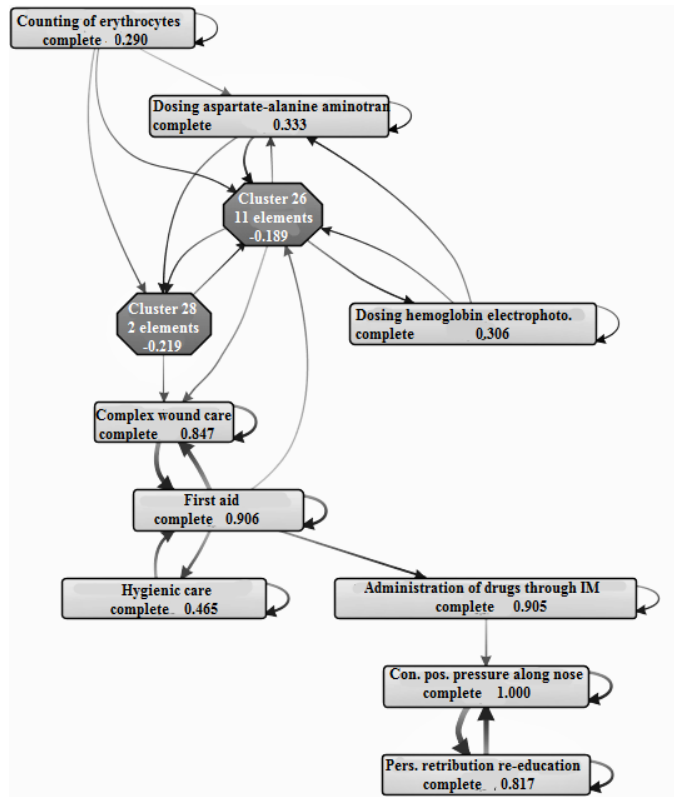


Figure 6: Fuzzy miner map as an illustration of the frequent occurrence of wound care and first aid activities, and dosing and counting activities. These two types of activities are connected through various testing activities. (This is a single column fitting image)

Tables

Table 1: Use cases extracted from previous research

Use case	Description	Limitations
Log preparation		
Dealing with events on the same day [29]	For events with a too low time granularity. The goal of this algorithm is to reduce the number of events happening at the same time. This will help to present processes correctly and deal with the problem of “pattern explosion”. This method is implemented as part of a BPI platform.	For the detection of clinical event packages expert knowledge is needed. The algorithm was developed within the perspective of clinical pathways, which are often hospital-specific and must be pre-defined.
Sequentializing events [30]	Suriadi et al. [30] observe that in some cases the strict ordering of events is unimportant. As such, the process model will show these events as “a series of sequential activities”.	This technique requires expert knowledge to determine in which cases the ordering of events is unimportant.
Abstraction & selection		
Pattern abstractions [18]	The technique includes pre-processing by means of abstraction in order to reduce the number of events and paths.	N.A.
Temporal abstraction [35]	Moskovitch and Shahar [35] applied data mining techniques on a diabetes case with unstructured processes and low abstraction levels.	It is in essence a data mining technique, which means that it does not take advantage of the dynamics of processes. Abstraction events would have to be defined beforehand since it is a classification tool and is meant to discover rules.
Activity Miner [19]	Günther and van der Aalst [19] propose a clustering technique to group low-level events from the log in the actual activities by recognizing an activity as a recurring pattern. There are three stages: initial scan; aggregation pass; final selection.	As the algorithm is based on the notion of proximity, the duration of activity executions must be significantly larger or smaller than the waiting time between distinct activities for a certain trace [19].
Drill-down and drill-up approach [31]	The drill-up approach looks at a log from the organizational perspective. It shows the path between departments instead of the path between activities. The drill-down approach looks at a particular sample of the dataset and compares the results of a mining plug-in with specialized knowledge.	Drill-up is a broad approach. The result can give you a view on the paths between departments. For more detail, on the providers or activities, one needs to drill down afterwards. The drill-down approach requires expert knowledge to perform data selection and rules or it requires guidelines to perform compliance checks.
ViTa-Lab framework [14]	This framework applies filtering mechanisms based on regular expressions. It is a stand-alone implementation.	N.A.

Clustering		
ActiTrac and the application of data and text mining [21]	The authors present a novel approach for trace clustering. They apply a separate cluster for atypical behavior. To deal with this cluster they propose a combination of data and text mining on process and case data.	There is a need for clearly prescribed processes and a limited amount of deviations. End goals need to be defined in advance in order to apply the data mining step. Text details are necessary to apply text mining.
Sequence clustering [36]	Clustering is based on the sequences of the traces using first-order Markov chains. The plug-in also contains several features for pre-processing.	The technique requires a limited amount of trace variants.
Process mining		
Fuzzy Miner [24]	This technique creates a process map taking into consideration significance and correlation of edges and nodes. The process maps automatically clusters activities. Only the most significant activities are shown.	Even with fuzzy miner incomprehensible models can arise.
Discovering data-aware declarative process models [37]	A combination of case and process data is used to predict events. The goal is to discover conditions on rules.	The technique requires a sufficiently large time span.
Social network discovery [27]	This technique mines social networks based on ‘provider’ information.	N.A.
Role Hierarchy miner [[25] in [26]]	This technique provides more information about which resources perform which activities.	The comprehensibility of the produced graph could be improved.
SCP-miner [33]	This technique deals with diversity of traces looking at the most frequent behavior as clinical pathways. The approach can be applied iteratively. The approach is implemented in C#.	All events are assumed to be clinical events, which are often hospital-specific. Moreover, it requires a complete pathway for patients.
Decision miner [38]	This approach analyses how data attributes influence the choices made in the process based on part process executions. Decision point analysis discovers which properties of a case might lead to certain paths in the process.	The technique requires a limited set of trace variants.
Dotted chart analysis [23]	The dotted chart is a fast tool to visualize the spread of events.	The plug-in has no filtering abilities, so filtering needs to be done beforehand.
Scatter diagram [14]	This diagram represents visually specific measurements of patients on a relative time scale. This is extended by maximal, minimal and means values of patients. The technique was implemented in the ViTa-Lab framework.	N.A.