A Process Mining approach to statistical analysis: application to a real-world advanced melanoma dataset

Erica Tavazzi[1,2], Camille L. Gerard[2], Olivier Michielin[2,3], Alexandre Wicky[2], Roberto Gatta[2], and Michel A. Cuendet[2,3,4]

[1] Department of Information Engineering, University of Padova, Italy
erica.tavazzi@phd.unipd.it
[2] Precision Oncology Center, Lausanne University Hospital (CHUV), Switzerland
[3] Molecular Modeling Group, Swiss Institute of Bioinformatics, Lausanne, Switzerland
[4] Department of Physiology and Biophysics, Weill Cornell Medicine, New York, USA

Abstract. Thanks to its ability to offer a time-oriented perspective on the clinical events that define the patient's path of care, Process Mining (PM) is assuming an emerging role in clinical data analytics. PM's ability to exploit time-series data and to build processes without any *a priori* knowledge suggests interesting synergies with the most common statistical analyses in healthcare, in particular survival analysis. In this work we demonstrate contributions of our process-oriented approach in analyzing a real-world retrospective dataset of patients treated for advanced melanoma at the Lausanne University Hospital. Addressing the clinical questions raised by our oncologists, we integrated PM in almost all the steps of a common statistical analysis. We show: (1) how PM can be leveraged to improve the quality of the data (data cleaning/preprocessing), (2) how PM can provide efficient data visualizations that support and/or suggest clinical hypotheses, also allowing to check the consistency between real and expected processes (descriptive statistics), and (3) how PM can assist in querying or re-expressing the data in terms of pre-defined reference workflows for testing survival differences among sub-cohorts (statistical inference). We exploit a rich set of PM tools for querying the event logs, inspecting the processes using statistical hypothesis testing, and performing conformance checking analyses to identify patterns in patient clinical paths and study the effects of different treatment sequences in our cohort.

Keywords: Process mining · Oncology · Melanoma · Statistical analysis

1 Introduction

Process Mining (PM) is a family of process analysis methods that aim at discovering, monitoring and improving the efficiency of real processes by extracting knowledge from the Event Logs (EL) recorded by an information system. Analytic algorithms are applied to ELs with the main goals of: (i) mining the data

in order to represent the process able to produce them (*Process Discovery*, PD), (ii) measuring to which extent a given process can represent an input EL or how much an EL complies with a given process (*Conformance Checking*, CC), and (iii) improving process efficiency, by allowing problem diagnosis and delay prediction, recommending process redesigns or supporting decision making (*Process Enhancement*) [2].

In PM for Healthcare (PM4HC), processes are meant as a graph of activities which can be performed with the aim of diagnosing, treating and/or preventing diseases to improve the patients' health status. The activities can be clinical and non-clinical and may represent different behaviours according to the specific organization [12]. Often, such processes are highly dynamic, complex, increasingly multidisciplinary [8]. Notably, the complexity increased recently due to the advent of personalized approaches to care, in which treatments are tailored to the specific profile of the patient and disease, such that the diversity of therapeutic pathways exploded compared to traditional standardized care guidelines.

Pragmatically, PM4HC has shown interesting applications in many domains, and in Oncology in particular, PM4HC was successfully applied to identify the most common patterns of care for many kinds of tumors, even though the purpose remained exploratory. Rectal cancer [7], gynecological cancer [11], and melanoma [13] were investigated both in terms of PD and CC, even if in most cases the focus was more on CC, while the application of PD remained descriptive of the general trend [9]. From this perspective, there were only few cases where the PM4HC analysis was used for statistical inference, *i.e.* to concretely develop predictive models assessing the role of covariates in determining disease evolution or patient clinical pathway. While the idea of applying a combination of PM and statistics for a complete statistical analysis is not entirely new [4][10], it is not a very common approach and still requires to be consolidated, in particular to integrate survival analysis, which plays a forefront role in Oncology.

In this work, we focus on exploring the contributions of PM when performing statistical analyses in Oncology. As an application, we examined a real-world cohort of advanced melanoma patients treated at the Lausanne University Hospital (CHUV); here we show how PM can guide and/or assist researchers in all the classical steps of statistical analysis, that is, data preprocessing, descriptive statistics, and inferential statistics. Figure 1 summarizes these steps.

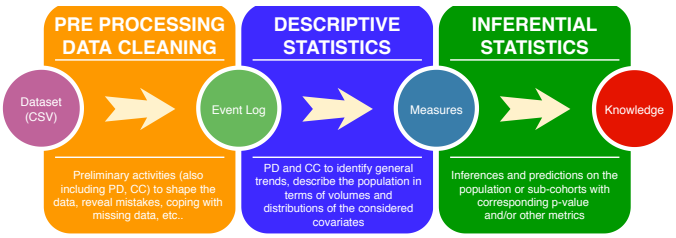


Fig. 1. Workflow of the classical steps of a statistical analysis, here implemented exploiting a process-oriented approach.

In the preprocessing step, we approached the data inspecting their structure, their information content, and their quality: after identifying the clinical milestones of interest (like diagnosis, treatments, survival outcome), data were first shaped as EL. We then employed the visualization tools provided by PM to detect data inconsistencies due to input errors or missing values. This allowed us to go back to the data sources, recheck and correct the recorded information, thus recursively improving the data quality.

In the descriptive analysis step, we first employed the EL time-oriented structure to inspect cardinality and order of the administered pharmacological treatments. Then, we implemented both unsupervised and supervised methods to capture the flow of the patients' pathways over data-driven graphs (PD approach) or user-defined graphs (CC approach), respectively. In this part of the analysis, the graphical output provided by PM allows a fast access to the design and/or interpretation of the models, and an immediate assessment of the treatments in terms of type, order and timing of consecutive administrations.

Finally, in the inferential statistics step, we build upon the processes constructed in the previous step to quickly select sub-cohorts of patients characterized by similar patterns of care and/or clinical attributes. The cohorts were then compared in terms of time-to-event outcome and overall survival (OS), using Kaplan-Meier analysis and log-rank test.

2 Material and Methods

2.1 Material

In this work, we analyzed the data of a cohort of patients treated at the CHUV and diagnosed with advanced melanoma.

Melanoma is an aggressive cancer that arises from melanocytes (pigment cells). Cutaneous melanoma is the most common type. However, it exists also uveal and mucosal melanomas, which occur in the eye and in the mucosa (such as the mouth or the vulva), respectively. The primary risk factor of cutaneous melanoma is ultraviolet light exposure. As outdoor activities are a way of life in Switzerland, the melanoma incidence is high in the country [3]. The extent of the disease progression is described by a staging system, ranging from I to IV: Stage IV indicates metastatization of melanoma cells to distant organs. Surgery is the most common and resolutive approach for the lowest stages, but when the disease is more extensive, systemic treatments such as Immunotherapy are required, with Radiotherapy also used as palliative or local treatment.

The study cohort includes 184 patients diagnosed with advanced melanoma between March 18th, 2008 and November 17th, 2019, with follow-up up to 2019, December 30th.[1] Data were sourced from the electronic healthcare records available at CHUV and curated by trained oncologists.

[1] This study was approved by the Research Ethical Committee of Canton de Vaud (CER-VD) and includes only patients who did not oppose usage of their data, and was conducted according to the Swiss Federal Act on Research involving Human Beings.

Data includes: sex, date of birth, primary tumor type, stage and diagnosis date, advanced tumor diagnosis date and mutation type (among BRAF-V600, BRAF-nonV600, NRAS, wild type (wt)), pharmacological treatments, and survival information (date of death or last follow-up). In this study, only the medications administered after the stage IV diagnosis were considered.

2.2 Methods

We implemented the classical statistical analysis pipeline shown in Figure 1 by employing PM4HC techniques to achieve the goals of each step. To perform the analyses, we used pMineR, an open source R library implementing PM4HC functionalities [5]. By handling data in the form of EL, it allows, among its features, to implement PD and CC analyses.

We started with the raw data set, which we first assumed to be *clean* from mistakes. First, we cast the data in the form of EL, by selecting the main clinical milestones of interest for the analysis and defining the rules to cope with missing values. Then, we implemented a PD algorithm based on First Order Markov Models (FOMMs)[5], to provide a fast and easy-to-understand representation of the subsequent events. This representation allowed us to identify visually some unexpected links between clinical events (*e.g.* due to mistakes in some dates). With the help of a physician, we iteratively reviewed the data and rerun the PD algorithm in order to increasingly approach the expected graph and thus refine the data quality.

To describe the general statistics of the population and quantify the flux of patients though different patterns of cares (the second step in Figure 1), we exploited both PD and CC techniques. The unsupervised PD analysis is based on the same FOMM model as described above. The supervised CC approach is based on a pre-defined representation of the different treatment lines implemented with the Pseudo-Workflow formalism (PWF) available in the software tool. Once performed PD and CC, the patients were grouped according to their paths through the graphs using the selection language provided by the tool. Then Kaplan-Meier survival curves and log-rank tests were used to quantify statistical differences between the groups, considering as end-points time-to-event in PD and OS in CC.

Process Discovery In PD, one of the most diffused process representation exploits the directly-follows graphs (DFGs): in this graphical representation, directed edges link all the couples of nodes representing subsequent activities in the EL. Even if DFGs have some well-known limitations [1], they are very intuitive and can be helpful to share with clinicians a first representation of the data. In the pMineR implementation, DFGs correspond to FOMMs.

Conformance Checking CC was performed by using the PWF, designing a diagram that describes the expected flow of events in terms of diagnoses, treatment lines, and survival events. Graphically, this results in a set of nodes,

representing the *status* that the subjects can assume, and a set of conditions (*triggers*) which fire transitions between status [6]. This representation allows to count which triggers/status are activated while automatically running down the events of each subjects, thus capturing the population behaviours through the diagram.

3 Results

3.1 Data preprocessing

Event Log For each patient, we built the EL with the following events, each associated with a time stamp:

- *Primary Stage*: the primary diagnosis, with melanoma type, tumor stage at the diagnosis, and somatic mutation harboured by the tumor as attributes;
- *Stage IV*: the diagnosis of stage IV;
- *T-Begin*: the begin of a line of treatment, with the type of the given drug(s) as attribute;
- *T-End*: the end of a line of treatment, with the type of the given drug(s) as attribute;
- *Dead, Censored*: the survival information, consisting in the dead of the patient or in the last follow-up date, respectively.

The collected treatments belong to the following categories:

- *Immunotherapy (IO)*: anti-CTLA4, anti-PD1, anti-CTLA4 + anti-PD1 (in combination), or other IO;
- *Chemotherapy (Chemo)*;
- *Targeted therapy*: tyrosine kinase inhibitors (TKI), other targeted therapy (TT).

In this study, only the treatments after stage IV diagnosis were considered.

Missing data In time-oriented analyses, missing information can consist either in unrecorded events or in missing dates associated to the events themselves. In order to preserve the clinical information we kept only complete treatments lines: the EL of patients with an incomplete line were thus truncated to the last available certain information (stage IV diagnosis or end of a previous line), artificially introducing a *Censored* event before the line with missing information.

Data Cleaning To detect mistakes in the data, we adopted an iterative approach: a FOMM process was discovered and visually analyzed to detect inconsistencies on unexpected edges. Then, the data were updated and the the procedure repeated until no more mistakes were found.

To give a practical example of detection, we report in Figure 2 a) the FOMM resulting from an intermediate version of the dataset, where unexpected edges emerge because the beginning of the first line of treatment was erroneously dated

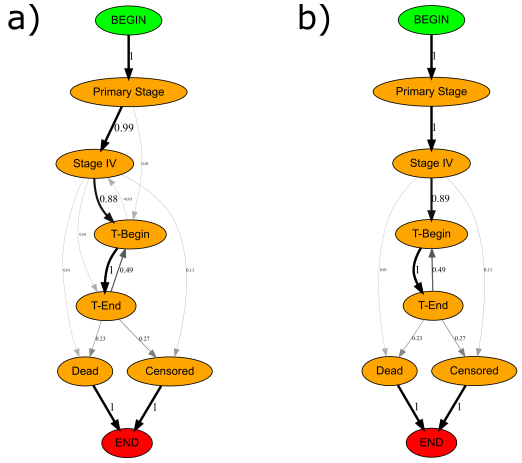


Fig. 2. First Order Markov Models obtained on all the events constituting the EL: a) before cleaning the information of a subject with an error in the dates, b) after data cleaning.

before the stage IV diagnosis for one patient in the source data. In Figure 2b) we can observe the FOMM after correction of the inaccurately collected information. This updated graph presents, conversely, only relations fully compliant with the nature (and the collection design) of the data.

With this approach we revealed some previously uncaught mistakes in the original data, such as inconsistency in data representation (*e.g.* dd/mm/yy vs dd/mm/yyyy), or temporal event inversion (*e.g.* cancer treatment begin before a tumor diagnosis).

3.2 Descriptive statistics

A first descriptive statistics was performed by querying the input EL, consisting of 1196 records: this allowed us to explore in the first instance cardinality and order of the administered treatments. Then, we delved into the data by using the FOMM, to obtain an agnostic data representation, and a PWF diagram, to verify the consistency of the process with respect to the expected behaviour.

Event Log querying By analysing the EL it was possible to perform some first descriptive investigations. We focused, specifically, on the treatments administered to the patients. Considering the events of all the patients, regardless of the position in the path of care, we extracted a total of 322 administered treatments. Table 1 reports, for each treatment category, its absolute and relative frequency of occurrence, and its duration in terms of median and inter-quartile range (25%-75%).

Out of 163 patients that received at least one recorded line of treatment, we identified 49 distinct patterns of treatment sequence. The most frequent ones are reported in Table 2.

Table 1. Occurrences and duration (in days) of the administered treatments collected in the data. The inter-quartile ranges (IQR) are computed at 25% and 75%.

Drug category	**Occurrences (%) (n=322)**		**Median (IQR) duration [days]**
TKI	76	(23.6)	122 (76.5–228.0)
anti-CTLA4 + anti-PD1	70	(21.7)	46.5 (0.0–167.8)
anti-PD1	66	(20.5)	84.0 (33.0–253.2)
anti-CTLA4	66	(20.5)	61.5 (31.0–63.0)
Chemo	29	(9.0)	44.0 (22.0–67.0)
Other IO	13	(4.0)	92.0 (22.0–203.0)
TT	2	(0.6)	461.5 (300.7–622.2)

Table 2. Most frequent patterns of treatment recorded in the data. The relative frequency of occurrence is computed over the total number of patients with at least one recorded treatment.

First line	**Second line**	**Occurrence (%) (n=163)**	
anti-CTLA4 + anti-PD1	-	36	(22.1)
anti-PD1	-	22	(13.5)
anti-CTLA4	-	11	(6.7)
anti-CTLA4 + anti-PD1	TKI	11	(6.7)
Chemo	anti-CTLA4	9	(5.5)
anti-CTLA4	anti-PD1	8	(4.9)
TKI	anti-CTLA4	6	(3.7)
anti-CTLA4	TKI	5	(3.1)
TKI	-	3	(1.8)

Process discovery on treatment sequences Figure 3 shows the FOMM obtained from the clean EL considering only the administered treatments (ignoring diagnosis and survival events). Such a process allows to inspect the temporal causality of the treatments, highlighting the most frequent connections over all the population. It also provides a first overview of the position of the treatments in the paths.

Conformance checking for treatment sequences We designed a PWF able to capture the chronological order of the events: at the top, we represented the events related to the staging, and then the different treatment lines. In order to be able to define treatments paths at different levels of granularity we added a further status for each treatment line, that is, *IO* (immunotherapy). This is doable thanks to the possibility in the PWF formalism to define simultaneous activation of multiple status. Finally, we introduced two additional status to catch the survival outcomes, namely *Dead* and *Censored*, that can be activated without constraints on the previous status, as soon as a survival event is read in the EL. The activation of the survival status terminates the inspection of the flow of events for that patient.

Figure 4 reports the result of the run on our cohort. Nodes and boxes report the number of times that a status/trigger was reached/fired. Due to space constraints, we limited the plot to the first two lines of treatment, even if the PWF included all the 7 lines of treatments available in the data.

By inspecting the graph, it is possible to follow the population's paths and read the corresponding number of subjects that run specific patterns. For in-

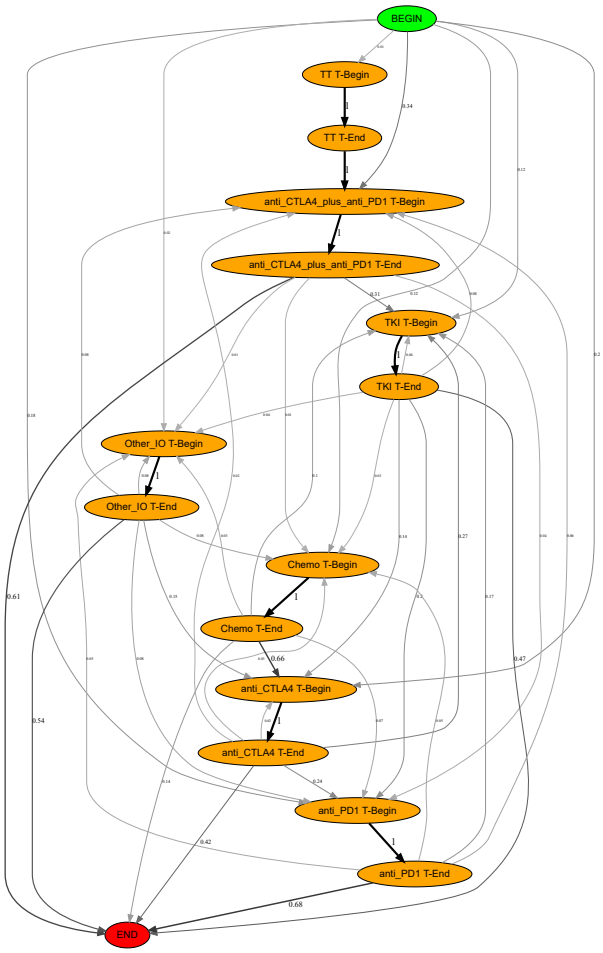


Fig. 3. First Order Markov Models obtained on the treatments.

stance, we can observe that all the patients included in the dataset (and thus with a BEGIN event) had a Stage IV diagnosis (expected by design), that the most frequent first line of treatment was the combination of anti-CTLA4 and anti-PD1 with a total of 56 occurrences, or that only 163 over 184 patients had a first line recorded, followed in 89 cases by a second line.

The survival nodes (*Dead* and *Censored*) are graphically separated from the others in order to limit the number of edges in the graph. However they can be reached from any point in the graph, and the available query tool can inspect at what precise point they were activated.

3.3 Inferential statistics

By exploiting the EL, the FOMM and the PWF diagrams of the previous analyses, we could easily select cohorts characterized by specific patterns of interest

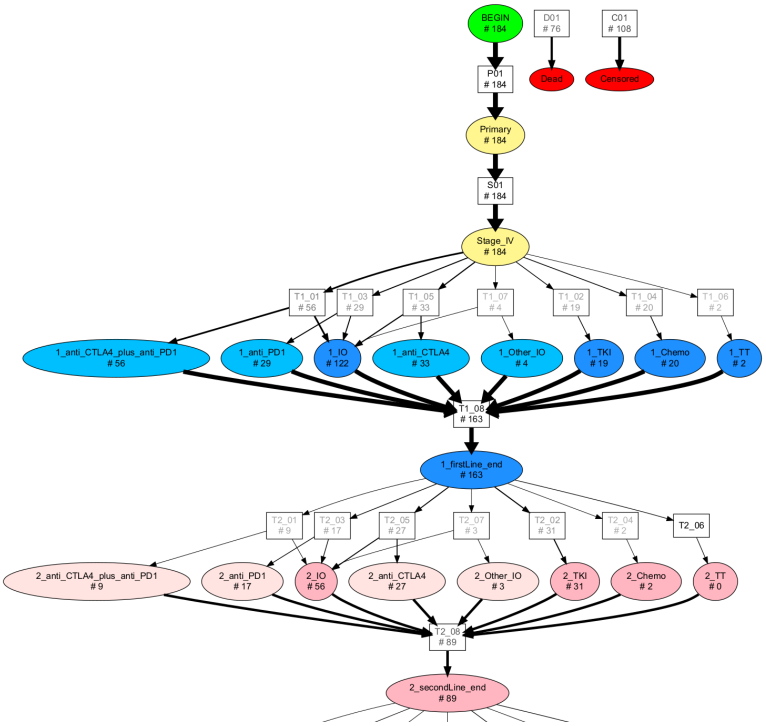


Fig. 4. Conformance Checking model (limited to the first two lines of treatments) reporting the status activated by the patients' processes over the used-defined PWF.

and perform survival analyses. While the FOMM strongly reflects (and is limited to) the events and the information present in the EL, the PWF represents an abstraction where the user has the opportunity to provide additional knowledge in the definition of the PWF structure itself. This enhanced semantic expressiveness is one of the main reasons why PWF was previously used in structuring Clinical Guidelines [5]. Descriptive statistics can help in suggesting hypotheses: in our case, the previous PWF and FOMM diagrams allowed to easily identify and query cohorts for statistical inference analyses. We report below two examples of the investigations we performed.

First, we inspected the relationship between type of somatic tumor mutation and time between primary and Stage IV diagnosis. Here, we consider the following mutation status: BRAF V600 mutated, BRAF non-V600 mutated, NRAS mutated, and wt. For this study, we limited the cohort to cutaneous melanoma patients, exploiting filtering tool to easily query the EL attributes.

We implemented a survival analysis by first using the FOMM structure of Figure 2 to query the path of interest (between the nodes Primary Stage and Stage IV) and obtain the time between the two events. Then, the Kaplan-Meier estimator is computed, with patients stratified by mutation status, as shown in Figure 5a). Even if a difference between the BRAF v600 mutated and the NRAS

mutated sub-cohorts seems to emerge, the log-rank test computed between all the survival distributions pairs report no significant differences (all p-values were >0.05) for any combinations.

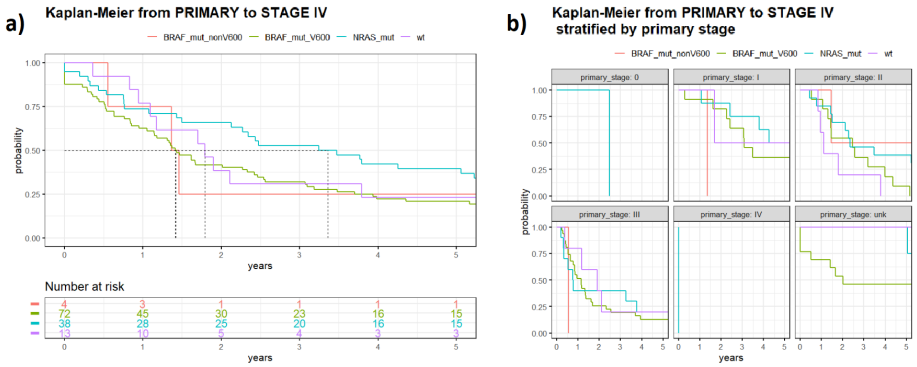


Fig. 5. Time-to-event analysis based on a mined FOMM: time from primary to stage IV diagnosis, stratified by: a) mutation, b) mutation and type of primary.

To demonstrate the potential of the analysis – even if in this case limited by the sample cardinality – we performed a further stratification of the data, distinguishing patients by their primary stage. Also here, pMineR facilitates this step, by allowing direct selection on the patient attributes. Figure 5b) reports the plot of the corresponding Kaplan-Meier estimator. Even if, as expected, no statistically significant clinical evidence emerges from this analysis, mainly due to the low number of subjects per category, it is interesting to observe how rapidly this approach allows to enrich the analysis' level of detail.

The second survival analysis exploits the PWF defined in Figure 4. We queried the data in order to identify any differences in terms of OS based on the following patterns of interest: (1) only IO (any BRAF status), (2) IO $\rightarrow$ TKI, (3) TKI $\rightarrow$ IO, (4) only TKI. In defining the rules, we grouped together consecutive lines belonging to the same category. Patterns interspersed with TT or Chemo treatments were excluded. Upon the suggestions of clinicians, in case of sequences with multiple treatment lines, only the first occurring pattern was considered. The resulting OS survival curves are shown in Figure 6. Table 3 reports the frequency of occurrence of each pattern, the median OS time (in years), and the percentage of patients alive at 1.5 and 3 years (CI at 95%), respectively. Statistical significance of OS differences was assessed with the log-rank test, which turned out to be significant for IO vs IO $\rightarrow$ TKI (p-value<0.0001) and IO vs TKI $\rightarrow$ IO (p-value: 0.012). The difference between IO and IO $\rightarrow$ TKI is expected because patients who receive TKI after IO are those who did not respond to IO. Knowing that the benefits of TKI are usually only temporary, it is not surprising that these patients have shorter OS. The difference between IO and TKI $\rightarrow$ IO is interesting, as it may be related to recent biological findings showing that acquired resistance to TKI may hinder IO efficacy.

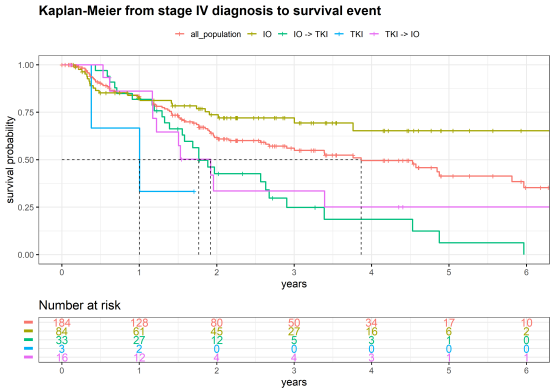


Fig. 6. Overall survival analysis based on a CC graph: time from stage IV diagnosis to death, stratified by treatment pattern.

Table 3. OS for the main treatment patterns of interest.

treatment path	**frequency**	**median OS [years]**	**1.5-year OS % (95% CI)**	**3-year OS % (95% CI)**
all	100 %	3.87	72.7 (66.1 - 80.1)	54.9 (47.1 - 64.1)
IO	45.7 %	NA	76.9 (68.0 - 86.9)	69.4 (59.1 - 81.5)
IO → TKI	17.9 %	1.77	63 (48.3 - 82.1)	18.6 (7.7 - 45.2)
TKI → IO	8.7 %	1.92	57.4 (36.6 - 90.1)	25.1 (9.7 - 65.3)
TKI	1.6 %	1.00	0	0

4 Discussion and Conclusion

PM4HC is expected to have an increasingly relevant role in the analysis of healthcare data, in particular in Oncology. Process-oriented representations, together with tools able to interrogate the data in terms of temporal patterns identified through paths in a workflow, are efficient ways to easily generate clinically-relevant hypotheses and measure statistical significance, in particular in survival analysis.

In this preliminary work, we demonstrated the added value of a process-oriented approach when performing three classical steps of data analysis: preprocessing, descriptive statistics, and inferential statistics. The main remarkable points emerging from this experience are: (a) query languages for EL, PD and CC are efficient tools for data cleaning and preprocessing, by quickly identifying previously unrecognized mistakes; (b) graphical representations can promote dialogue between clinicians and data scientists, suggesting alternative perspectives and possible research questions; (c) PD gives a relevant contribute in representing the data in an agnostic way; on the other hand CC (with formalisms such as PWF) allows implementing multi-scale data abstractions and identifying patterns or inconsistencies of the data in pre-defined workflows; (d) the process representations, both in PD and CC, effectively support survival analysis techniques, allowing rapid definition of sub-cohorts of interest and providing immediate statistical measures of differences between various paths of the graph.

Noticeably, each step of this study was performed in close cooperation between clinicians and PM scientists, in the effort of creating a multidisciplinary team with shared PM skills. The final goal will be to give full autonomy to physicians to perform PM analyses themselves.

In the future, PM4HC has great potential to be developed further in synergy with classical statistical tools to analyze healthcare-related data. In particular, the fast-growing amount of real-world clinical data produced in modern hospitals, each patient's therapeutic journey being by nature a temporal process, represents a formidable opportunity for PM4HC to contribute to the advent of precision medicine.

References

1. van der Aalst, W.: A practitioner's guide to process mining: Limitations of the directly-follows graph. Procedia Computer Science **164**, 321–328 (2019)
2. van der Aalst, W., Adriansyah, A., et al.: Process mining manifesto. In: Int. Conf. on Business Process Management. pp. 169–194. Springer (2011)
3. Bulliard, J., Panizzon, R., Levi, F.: Melanoma prevention in Switzerland: where do we stand? Revue medicale suisse **2**(63), 1122–1125 (2006)
4. Cowey, C.L., Liu, F.X., Boyd, M., Aguilar, K.M., Krepler, C.: Real-world treatment patterns and clinical outcomes among patients with advanced melanoma: A retrospective, community oncology-based cohort study (A STROBE-compliant article). Medicine (Baltimore) **98**(28), e16328 (Jul 2019)
5. Gatta, R., Lenkowicz, J., et al.: pMineR: an innovative R library for performing process mining in medicine. In: Conf. on Artificial Intelligence in Medicine in Europe. pp. 351–355. Springer (2017)
6. Gatta, R., Vallati, M., Lenkowicz, J., et al.: Generating and comparing knowledge graphs of medical processes using pminer. In: Proceedings of the Knowledge Capture Conference. K-CAP 2017, Association for Computing Machinery, New York, NY, USA (2017)
7. Geleijnse, G., Aklecha, H., et al.: Using process mining to evaluate colon cancer guideline adherence with cancer registry data: a case study. In: AMIA (2018)
8. Homayounfar, P.: Process mining challenges in hospital information systems. In: 2012 Federated Conf. on Computer Science and Information Systems. pp. 1135–1140. IEEE (2012)
9. Kurniati, A.P., Johnson, O., Hogg, D., Hall, G.: Process mining in oncology: A literature review. In: 2016 6th Int. Conf. on Information Communication and Management. pp. 291–297. IEEE (2016)
10. Lenkowicz, J., Gatta, R., et al.: Assessing the conformity to clinical guidelines in oncology: An example for the multidisciplinary management of locally advanced colorectal cancer treatment. Management Decision **56**(10), 2172–2186 (Oct 2018)
11. Mans, R., Schonenberg, H., Song, M., Aalst, W.V., Bakker, P.: Application of process mining in healthcare - a case study in a dutch hospital. In: BIOSTEC (2008)
12. Mans, R.S., van der Aalst, W.M., Vanwersch, R.J.: Process mining in healthcare: evaluating and exploiting operational healthcare processes. Springer (2015)
13. Rinner, C., Helm, E., Dunkl, R., et al.: Process Mining and Conformance Checking of Long Running Processes in the Context of Melanoma Surveillance. Int J Environ Res Public Health **15**(12) (12 2018)

Process Mining of Disease Trajectories in MIMIC-III: A Case Study

Guntur Kusuma[1,3,*[0000-0002-0208-125X]], Angelina Kurniati[2[0000-0002-4747-1067]], Ciarán D. McInerney[1[0000-0001-7620-7110]], Marlous Hall[4[0000-0003-1246-2627]], Chris P. Gale[4[0000-0003-4732-382X]], Owen Johnson[1[0000-0003-3998-541X]]

[1] School of Computing, University of Leeds, UK LS2 9JT
[2] School of Computing, Telkom University, Bandung, Indonesia 40257
[3] School of Applied Science, Telkom University, Bandung, Indonesia 40257
[4] Leeds Institute of Cardiovascular and Metabolic Medicine, University of Leeds, UK LS2 9JT
*scgpk@leeds.ac.uk

Abstract. A temporal disease trajectory describes the sequence of diseases that a patient has experienced over time. Electronic health records (EHRs) that contain coded disease diagnoses can be mined to find common and unusual disease trajectories that have the potential to generate clinically valuable insights into the relationship between diseases. Disease trajectories are typically identified by a sequence of timestamped diagnostic codes very similar to the event logs of timestamped activities used in process mining, and we believe disease trajectory models can be produced using process mining tools and techniques. We explored this through a case study using sequences of timestamped diagnostic codes from the publicly available MIMIC-III database of de-identified EHR data. In this paper, we present an approach that recognised the unique nature of disease trajectory models based on sequenced pairs of diagnostic codes tested for directionality. To promote reuse, we developed a set of event log transformations that mine disease trajectories from an EHR using standard process mining tools. Our method was able to produce effective and clinically relevant disease trajectory models from MIMIC-III, and the method demonstrates the feasibility of applying process mining to disease trajectory modelling.

Keywords: Disease trajectories, Process mining, Electronic Health Records.

1 Introduction

There is a small but growing body of literature exploring the generation of disease trajectories using electronic health records (EHR) [1, 2]. The rich collection of patient data in the EHR is a valuable source to get an extensive trail of disease diagnoses over time [3]. Mining the trails of disease diagnoses and the temporal information may help to identify patterns in disease trajectories of clinical value. A better understanding of patterns of disease may advance precision medicine to improve care at an individual level [4] and improve medical understanding of common disease progression at the population level [5, 6]. A study by Jensen et al. [7] had identified the disease trajectories of a

large cohort by combining a data-driven and statistical approach. However, their trajectories were built based on overlapping pairs of diagnostic codes suggesting the presence of longer trajectories without confirming if such trajectories are available in the data. Based on this, we propose an improvement by incorporating process mining as a toolset and method for mining end-to-end disease trajectories.

Process mining utilises a set of tools to discover process models using data from an organisation's information system. Extracted data are transformed into an event log, a collection of activities and its corresponding timestamps, sometimes supplemented with additional attributes. There is now a large body of literature applying process mining to the domain of healthcare, typically focussed on discovery of actual care processes [8], conformance to guidelines and enhancement to improve the quality of healthcare services [9], the safety of the patients, and better management of resources [10, 11].

Jensen et al. [7] defined a disease trajectory as the patient's orderly series of diagnoses. The definition is comparable to the concept of a trace in process mining where a trace is the sequence of activities for an individual case [12]. We hypothesise that it should be feasible to apply process mining to discover a disease trajectory model [2]. To the best of our knowledge, this is the first time process mining has been used to identify disease trajectories from a real world EHR.

In this paper, we present a novel disease trajectory mining method using process mining techniques applied to the MIMIC-III open access EHR database. We identified the sequence of diagnoses (trace) based on the temporal aspect of the patients' admissions, broke down each trace into pairs of diagnoses, statistically analysed the pair's correlation and represented the identified disease trajectories using a directly-followed graph produced by standard process mining visualisation tools [12]. The research questions are as follow: *Q1-Can disease trajectories be identified using a process-mining approach? Q2-What are the most followed trajectories and what exceptional trajectories are followed? Q3-Are there differences in trajectories followed by different patient groups (by sex, by age group, by mortality status)?* And, *Q4-What are the longest and shortest average time transition trajectories?*

2 Background

Process mining provides a set of techniques and tools to uncover the real behaviour of processes from a range of perspectives including, but not limited to [12]: control-flow, performance, conformance, and organisational. There are three types of process mining: first, process discovery to generate process models from event log data, second, process conformance to check either a process model conforms to an event log or vice versa and third, process enhancement to improve a process model using the information of the actual process recorded in the event log [12].

In healthcare, process mining techniques may help the clinicians answer questions associated to each characteristic of the healthcare processes (e.g. primary care, secondary care, tertiary care, etc.) [8]. The rich information in the EHR is the source of answer

to the four types of data science questions: "*what happened?*", "*why did it happen?*", "*what will happen?*", and "*what is the best that may happen?*". In this study, we followed the most widely used methodology, the PM^2 framework, which describes six process mining stages and defines the set of activities to complete each stage.

The diagnostic codes available within electronic health records result from diagnostic decisions made by clinical specialists after considering the patient's health problem [13]. Jutel [14] described the diagnosis as a process of assessing and making a formal judgement based on a specific physical symptom that takes place at a particular time involving both patient and doctor. Once the disease is determined it is recorded in the EHR using standard diagnostic codes such as the World Health Organisation's International Classification of Diseases (ICD) [15].

3 Method

The goal of this case study was to identify patients' disease trajectories using a process-mining approach. We conducted a retrospective cohort study of patients who were admitted to critical care using the MIMIC-III database as our data source [16]. The MIMIC-III database contains a detailed record of patients' clinical care that has been de-identified to respect the sensitive nature of the data. It is available online to researchers (https://mimic.physionet.org) under an open access policy. We obtained access through two mandatory steps: a training program in human research subject protections and a data user agreement. The Process Mining Project Methodology (PM^2) was followed in this study as the methodology allows us to have multiple research questions that require iterations of analyses [17].

3.1 Data source for the case study

MIMIC-III provides a database of de-identified electronic health records containing the medical history from 2001 to 2012 of 46,520 critical care patients extracted from the EHR of the Beth Israel Deaconess Medical Centre in Boston, USA [16]. The database includes data on patient demographics, laboratory tests, diagnostic codes (in ICD-9 coding standard), medications, bedside monitoring, clinicians' notes and reports, and death records (linked to Social Security Death Index for outpatient death). As part of the anonymisation process, the timestamps used in the MIMIC-III dataset have been intentionally shifted into the future (between 2100 and 2200) by a random offset generated for each patient. This means that the sequence of disease codes and the time intervals between disease codes has been preserved for individual patients but no comparisons between patients are possible. This does not affect disease trajectory mining, but does limit other process-mining approaches such as the identification of bottlenecks. Our group has experience of applying process mining to MIMIC-III and in earlier work have published a data quality assessment on the suitability of the various MIMIC-III data components that are compatible with process mining [18].

3.2 PM^2 for Disease Trajectory Mining

In this section, we identify those sections of the PM^2 that we have adapted for disease trajectory mining. For a full understanding of the PM^2 method see [17].

In Stage 1 (Planning), our research questions were identified from a literature review and confirmed by a project team composed of a clinician, and epidemiologist and process mining and data science researchers.

In Stage 2 (Extraction), we defined the scope by determining the granularity level of data, the time period, and attributes of interest. The MIMIC-III database contains admissions of adult patients aged 16 years old or older [16] who were admitted to the hospital between 1 June 2001 and 10 October 2012. Only patients with at least two admissions were selected to capture the progression of the disease. Patients were followed up for mortality status until the last available discharge as the last censoring date and time for those who died within the hospital. The censoring date for patients who died outside of the hospital is the date recorded in the social security master death index in the MIMIC-III database. We used the first 3-digit ICD-9 codes to indicate diagnoses, [19] but excluded codes known not to be related to development of diseases, e.g. administration codes. Event data were extracted from the ADMISSIONS, PATIENTS, and DIAGNOSES_ICD tables in MIMIC-III database as the input for creating an event log (**Table 1**). The time of admission was used as the activity timestamp and the diagnostic code as the activity name. The patients were grouped according to their age in bands of 5 years. The attribute of age group was calculated from the patient's age at first admission.

In Stage 3 (Data Processing), we created the event log as defined in the PM^2 by creating the views, then filtering and enriching them. The case identifier for each event was taken from the patient identifier (`subject_id`), the diagnostic code was used as the event name (`diagnosis_code`), and the admission time as the timestamp (`admittime`). The event log was filtered by removing recurring diagnostic codes (retaining the first occurrence), then reapplying the exclusion of patients with only one diagnostic code. The sequences of diagnostic codes for each patient in the event log informed a set of ordered pairs of diagnostic codes, D1→D2, where the diagnostic code D1 preceded the diagnostic code D2. For example, a patient's event log, D1→D2→D3, informed two ordered pairs of diagnostic codes, D1→D2 and D2→D3. We excluded ordered pairs that occurred only once. To measure the strength of association between the ordered pairs, we compared the probability of diagnosis D2 occurring among patients who did and did not have a D1 diagnosis previously in the event log. This relative risk (RR) [20] indicated whether the D2 diagnosis was more incident in the group with a D1 diagnosis (RR > 1), less incident in the group with a D1 diagnosis (RR < 1), or equivalent (RR = 1). The RR is calculated as

$$RR = \frac{(a/(a+b))}{(c/(c+d))} \tag{1}$$

where a is the number of patients having D1 and D2, b is the number of patients having D1 but not D2, c is the number of patients without having D1 but having D2, and d is the number of patients neither having D1 nor D2.

Table 1. Source of the required data from MIMIC-III database

Variables	Table source in MIMIC-III	Field name
Case identifier	PATIENTS	subject_id
Event	DIAGNOSES_ICD	hadm_id, icd9_code, seq_num
Activity name	DIAGNOSES_ICD	icd9_code (first 3 digits)
	ADMISSIONS	hospital_expire_flag
	PATIENTS	expire_flag (translated into 1:Dead, 0:End of data)
Time stamps	ADMISSIONS	admittime,dischtime, deathtime
	PATIENTS	dod, dod_hosp, dod_ssn,
Sex	PATIENTS	gender
Age*	PATIENTS	dob
	ADMISSIONS	admittime
Age group**	PATIENTS	dob
	ADMISSIONS	admittime

* the age calculation using PATIENT's dob and ADMISSIONS's admittime.

** the variable was added to group the patients' age.

Following Jensen et al [7], only pairs with RR > 1 were carried forward for further processing. For a given pair of diagnoses D1 and D2, it was possible for both D1→D2 and D2→D1 trajectories to satisfy the RR > 1 threshold. Our goal was to identify disease trajectories that were acyclic, so we carried forward the dominant directionality of a given pair of diagnostic codes, only. We applied one-tailed binomial tests [21] to define the dominant directionality of pairs, i.e. D1→D2 or D2→D1. Using a significance level of $\alpha = 0.05$, only ordered pairs of diagnostic codes with one statistically significant direction were carried forward to define the final pairlog.

The final pairlog was transformed back into an event log and recurring diagnoses in each trace were merged to avoid loops. The event log was then enriched by adding attributes of age at admission, sex, age group and the mortality status. These attributes were not used to define the disease trajectory models, but allowed post-hoc analyses to determine differences between disease trajectories according to each attribute. The enriched event log was then loaded into ProM, an open-source process mining tool (https://promtools.org). A START and END event was added to every case in the event log to provide common start and end points of traces. The final event log then converted into the XES format. Common traces were grouped in trace variants using the Explore Event Log (Trace Variants/ Searchable/ Sortable) feature in ProM [22].

In Stage 4 (Mining and Analysis) we used ProM to analysed the event log to identify unique trace variants, performed process discovery, visualised the discovered model and performed conformance checking. For process analysis, we calculated descriptive summary statistics of the disease trajectories that were identified, including stratification by patient groups. The event log was visualised using the Explore Event Log (Trace variants/ Searchable/ Sortable). The Interactive Data-aware Heuristics Miner (iDHM) [23] plug-in was used to discover the disease process models.

The quality of the discovered models were evaluated using replay fitness, precision and generalisation [24]. Replay fitness is a measure of how many traces from the log can be reproduced in the process model, with penalties for skips and insertions. Precision is a measure of how 'lean' the model is at representing traces from the log. Lower

values indicate superfluous structure in the model. Generalisation is a measure of generalisability as indicated by the redundancy of nodes in the model; The more redundant the nodes, the more variety of possible traces that can be represented. The value of each measure represents by a number between 0–1. Discovery and conformance checking used plugins in ProM. The Replay a Log on Petri Net for Conformance Analysis plug-in for measuring the fitness [25], Align-ETConformance plug-in [26] for the precision, and the Measure Precision/Generalization plugin for measuring the generalisation. Other tools used in this study were PostgreSQL as the database management system of MIMIC-III, and Python through Jupyter Notebook [27].

4 Results

An event log was extracted from an EHR to identify disease trajectories, pairs of diagnoses were identified and analysed for correlation measurement and tested for directionality. The discovery algorithm is applied to produce the disease trajectory model and represented using the directly-followed graph.

In Stage 1 (Planning), we aimed to mine the disease trajectory agnostically without any specific selection of diagnosis and time window. Following the literature review in section 2, we defined the main research question as: (*Q1) Can disease trajectories be identified using a process-mining approach?* Further questions added which were motivated by the frequently posed question for process mining in healthcare [28]: *(Q2) What are the most followed trajectories and what exceptional trajectories are followed?(Q3) Are there differences in trajectories followed by different patient groups (by sex, by age group, by mortality status)? (Q4) What are the longest and shortest average time transition trajectories?*

In Stage 2 (Extraction), Of the 58,976 unique admissions in MIMIC-III from 46,520 patients, there were 6,984 unique ICD-9 diagnostic codes used for 651,000 diagnoses. From this dataset, we excluded 172,685 (26.5%) diagnostic codes that are medically known to be codes related to external factors not directly related to the development of diseases [5], including pregnancy (ICD-9 3-digit codes 630-679, 760-779), general symptoms and signs not related to a disease (780-799), external cause (800-999, E800-E999), and administration (V01-V89). We further excluded 436,483 (67%) secondary diagnostic codes and focused on the 41,832 primary diagnostic codes whilst there will be valuable opportunity in exploring the secondary diagnostic codes.

In Stage 3 (Data Analysis), we composed the selected variables in a way that follows the minimum requirements of event log (see **Fig. 1.a**). The traces of each patients are illustrated in **Fig. 1.b**. We removed 2,692 (16.2%) recurrent diagnoses, retained the first occurrence, excluded patients with only one admission, and subsequently excluded patients who were less than 16 years old at their first ever admission. A total of 4,911 patients remained in the event log consisting of 11,725 diagnostic codes. **Fig. 1** shows the transformation of event logs into a log of ordered pairs of diagnostic codes (pairlog)(see **Fig. 1.c**). The resulting pairlog contained 6,814 ordered pairs of diagnostic codes. Only 3,781 pairs remained after filtering for RR > 1 and the binomial tests for

directionality suggested there were 826 ordered pairs of diagnostic codes with a statistically significant dominant direction. The resulting data contained 796 traces where each trace represents a patient's disease trajectory.

subject_id	diagnostic_code	timestamp
21	410	11/09/2134 12:17
21	038	30/01/2135 20:50
124	433	24/06/2160 21:25
124	441	17/12/2161 03:39
124	440	21/05/2165 21:02
124	569	31/12/2165 18:55

(a) The extracted event log

#21: 410→038

#124: 433→441→440→569

(b) The trace of diagnosis

subject_id	Antecedent	Subsequent	Time1	Time2
21	410	038	11/09/2134 12:17	30/01/2135 20:50
124	433	441	24/06/2160 21:25	17/12/2161 03:39
124	441	440	17/12/2161 03:39	21/05/2165 21:02
124	440	569	21/05/2165 21:02	31/12/2165 18:55

(c) The *pairlog*

Fig. 1. Illustration of the transformation steps of event log for pairwise analysis. (a) The extracted event log from MIMIC-III; (b) the illustration of traces of diagnoses for each patient; (c) the transformed event log into pairlog.

In the last step of filtering, we transformed the pairlog back to an event log and enriched with age at admission, sex, age group and the mortality status. We then loaded the enriched event log into ProM, artificial 'START' and 'END' events were added and then analysed the trace variants using the Explore Event Log feature. Among the 796 traces, we further removed twenty traces that were unique to a single, individual patients as part of good anonymisation practice. Finally, the 776 common traces found in the event log were grouped into 81 trace variants.

In Stage 4 (Mining and Analysis), there were eighty one unique trace variants informed the processing discovery algorithms to answer the Q1. The conformance of the discovered disease trajectory model demonstrated fitness = 0.93, precision = 0.94, and generalisation = 0.92. Further evaluation was done by 5-folds cross-validation where the original event log was randomly divided into five groups of sub-event log equally. One sub-event log was used as the validation data and the remaining four sub-event logs as training data. The cross-validation process was done five times to allow each sub-event log used once as the validation data. The average value from the cross-validation are expected to be lower than the conformance, resulting fitness = 0.92 (SD: 0.006), precision = 0.82 (SD: 0.06), and generalisation = 0.88 (SD: 0.02). This suggests that the discovered trajectory model (**Fig. 2**) is robust to sampling, allows the traces seen in the event log, is precise enough to not allow behaviour unrelated to what was seen in the event log, and general enough to reproduce future behaviour of the trajectories.

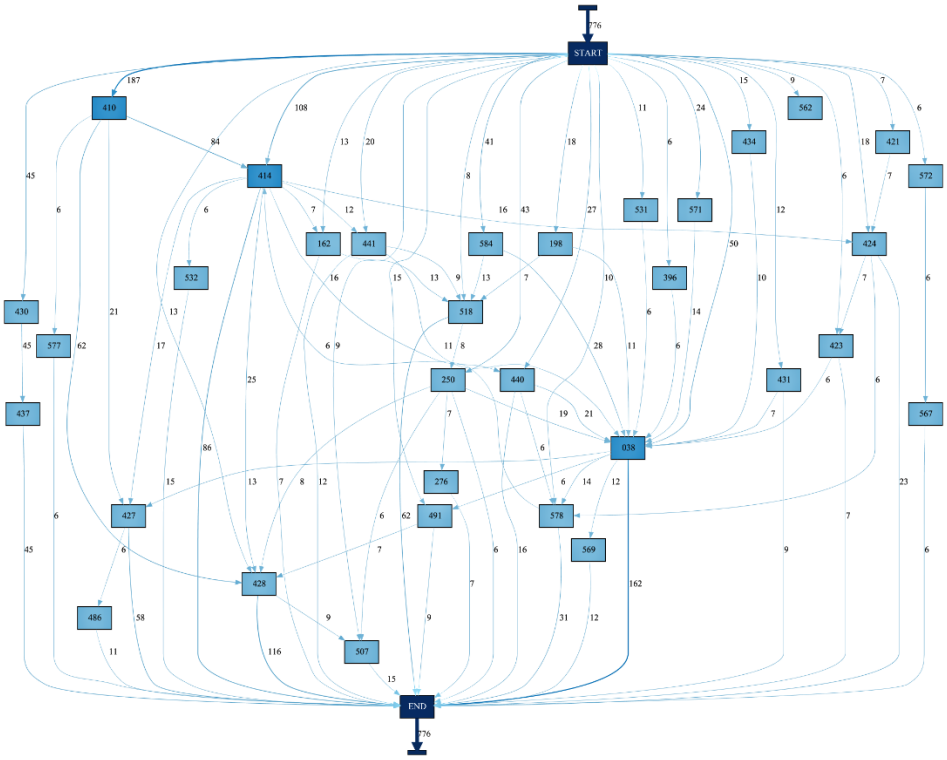


Fig. 2. The directly-follow graph representation of Disease Trajectory Model of Critical Care patients in MIMIC-III with the minimum case frequency = 6.

In respond to the Q2, among 776 patients there are 81 distinct trajectories (**Table 2**). The most-followed trajectory (n=80; 10.3%) was acute myocardial infarction to ischemic heart disease, which is consistent with the published literature [7, 29, 30]. Septicaemia occurred most frequently (n=212; 27.3%), both as a precedent (n=50; 6.4%) and subsequent (n=162; 20.9%), with mortality in the end (n=143; 66.9%). This supported previous findings that it is associated with morbidity and mortality [16, 31]. There are three exceptional trajectories of two patients each (0.26%) (**Table 2**).

Table 2. The three most-common and least-common trace variants.

Traces (%)	Trace Variant	Median (months)	Dead (%)	Male (%)
80 (10.31%)	START→410→414→END	6.5	75	70
62 (7.99%)	START→410→428→END	3.9	72.58	54.84
45 (5.80%)	START→430→437→END	3.9	4.44	35.56
…	…	…	…	…
2 (0.26%)	START→410→427→486→END	28.3	100	50
2 (0.26%)	START→507→491→482→END	43.6	50	100
2 (0.26%)	START→518→250→038→END	14.6	100	0

ICD-9 Codes translation: 038 = Septicaemia, 250 = Diabetes mellitus, 410 = Acute myocardial infarction, 414 = Ischemic heart disease, 427 = Cardiac dysrhythmias, 428 = Heart failure, 430 = Subarachnoid haemorrhage, 437 = Other and ill-defined cerebrovascular disease, 482 = Other bacterial pneumonia, 486 = Pneumonia, organism unspecified, 491 = Chronic bronchitis, 507 = Pneumonitis due to solids and liquids, 518 = Other diseases of lung.

The third question was (Q3) Are there differences in trajectories followed by different patient group? We answered the question by comparing trajectories by sex (male, female) and age band (18-34 years, 35-64 years, and >64 years). The male cohort consisted of 447 patients with the median duration of follow-up 6.98 months (IQR 1.6 – 28.2) where 252 cases (56.3%) ending in death. The most-common trajectory was acute myocardial infarction followed by other forms of chronic ischemic heart disease (56 cases, 12.5%) with median interval 6.5 months (IQR 1.5 – 35.3). In the female cohort, there were 329 patients with the median duration of follow-up 7 months (IQR 2 – 24.4) where 176 cases (54.4%) ending in death. The most-common trajectory was subarachnoid haemorrhage followed by other and ill-defined cerebrovascular disease,(29 cases, 8.8%) with median interval 3.4 months (IQR 2.3 – 7.5). The most-followed trajectory in a group of 18 to 34-year-old cohort was diabetes followed by hypertensive chronic kidney disease (3 cases) with median interval 55.8 months (IQR 33 – 56.5). For the group of 35 to 64 years, there were 44 cases (14.5%) with acute myocardial infarction followed by ischemic heart disease, with median interval 7.8 months (IQR 1.9 – 39.7). Among 329 cases in this age group, there were 133 cases (40.4%) ending in death. Patients in >64 years, there were 293 (68.1%) deaths while the most-common trajectory was acute myocardial infarction followed by heart failure, with median interval 4.7 months (IQR 1.5 – 21.8).

The fourth question was (Q4) What are the longest and shortest average time transition trajectories? The longest disease progression at 63 months was *Ischemic heart disease* to *Diverticula of intestine* while the shortest progression was *Gastrointestinal hemorrhage* to *Liver abscess and sequelae of chronic liver disease* with average time transition is less than a month (0.98) (**Table 3**).

Table 3. The three longest and shortest average time interval trajectories in MIMIC-III.

Antecedent	Subsequent	Mean*	Median (IQR)**
A. The three longest average time interval trajectories (descending)			
Chronic ischemic heart disease	Diverticula of intestine	63	75.9 (54 – 84.8)
Chronic ischemic heart disease	Occlusion of cerebral arteries	52.7	51.2 (40.4 – 52.6)
Chronic ischemic heart disease	Heart failure	46	41.5 (4.6 – 89.7)
B. The three shortest average time interval trajectories (ascending)			
Gastrointestinal hemorrhage	Liver abscess and sequelae of chronic liver disease	0.98	0.81 (0.6 – 1.3)
Other diseases of endocardium	Other diseases of pericardium	1	0.8 (0.6 – 1.13)
Chronic bronchitis	Other bacterial pneumonia	2.2	2.2 (1.6 – 2.7)

*Mean is in months. **Median is in months (IQR); IQR = interquartile range.

5 Discussion

We present a case study of 776 patient admissions associated with 81 different disease transitions to demonstrate the feasibility of using a process-mining approach to reveal disease trajectories using a hospital electronic health record database. We show that the PM^2 framework is suitable for mining disease trajectories and is complemented by the addition of descriptive summary statistics in Stage-3 (Data Processing). Our approach

applies a number of transformations to the data, which were adapted from published disease trajectory methods for constructing selected pairs of diagnoses with strong correlation, followed by testing the pairs' directionality to form the trajectories.

Process mining offers techniques to discover disease trajectories and measure the quality of the algorithm to discover the trajectory model. In this work we presented replay fitness, precision, generalisation and cross-validation to validate the model. The process-mining approach opens opportunities to cross-reference discovered disease trajectories with other critical care event data by defining workflows that can actioned using widely-available software. By conducting conformance checking, we have the indicators to show if the discovered model has a good quality. We note that the earlier study by Jensen et al. [7], did measure the robustness of their discovered disease trajectory model with one indicator that is similar to the replay fitness in process mining. This approach is useful to validate that the final model conforms closely to the data.

A particular benefit of the process-mining approach to constructing disease trajectories is that it may provide summaries of cases, events and time interval between occurrences of disease. For example, our method identified the trajectory of *acute kidney injury* (AKI) (584) followed by *septicaemia* (038) with an average interval of 16.22 months. This finding supports the conclusion of [32] where sepsis was a frequent consequence after AKI in intensive care setting. Also, the process-mining approach could provide an estimation of sepsis development after AKI as suggested in [33]. Our method also incorporates additional case attributes that easily facilitate outputs to be stratified by specific characteristics, e.g. sex, age group, and mortality status. For example, although the data were not pre-stratified for females, process mining tools made it easy to query the event log to reveal a dominant trajectory in females – *subarachnoid haemorrhage* (430) followed by *other and ill-defined cerebrovascular disease* (437) – that agrees with previous research [34].

6 Conclusion

In this paper, we have presented the mining of disease trajectories using a process-mining approach. The mining used the MIMIC-III dataset which is comparable to many databases from EHR systems in use at hospitals across the world. Our study included the use of PM^2 framework to mine a representative disease trajectory model from an EHR and addressed quality dimension standards. This study opens opportunities for future works in implementation of the technique using population sized EHR data. We believe the association of pairs of diagnoses might be improved by null hypothesis significance testing of relative risk rather than magnitude-based testing. Future work might assess the sensitivity of the method to the choice of process discovery algorithm used to mine the disease trajectory model.

7 Acknowledgements

The research was supported by the National Institute for Health Research (NIHR) Yorkshire and Humber Patient Safety Translational Research Centre (NIHR YH PSTRC) and the Indonesia Endowment Fund for Education (LPDP).

References

1. Allam, A., et al., *Patient Similarity Analysis with Longitudinal Health Data.* arXiv preprint arXiv:2005.06630, 2020.
2. Kusuma, G., et al. *Process Mining of Disease Trajectories: A Feasibility Study*. in *13th International Conference on Health Informatics*. 2020.
3. Weber, G.M., K.D. Mandl, and I.S. Kohane, *Finding the missing link for big biomedical data.* JAMA, 2014. **311**(24).
4. Jensen, P.B., L.J. Jensen, and S. Brunak, *Mining electronic health records: Towards better research applications and clinical care*. 2012. p. 395-405.
5. Hanauer, D.A. and N. Ramakrishnan, *Modeling temporal relationships in large scale clinical associations.* J Am Med Inform Assoc, 2013. **20**(2): p. 332-341.
6. Rothman, K.J. and S. Greenland, *Causation and causal inference in epidemiology*. American Journal of Public Health, 2005. **95**(SUPPL. 1): p. S144-50.
7. Jensen, A.B., et al., *Temporal disease trajectories condensed from population-wide registry data covering 6.2 million patients.* Nature Comm., 2014. **5**(May): p. 1-10.
8. Mans, R.S., W.M.P. van der Aalst, and R.J.B. Vanwersch, *Process Mining in Healthcare Evaluating and Exploiting Operational Healthcare Processes*. 1 ed. 2015: Springer International Publishing. 91.
9. Partington, A., et al., *Process Mining for Clinical Processes: A Comparative Analysis of Four Australian Hospitals.* ACM Trans. Manag. Inform. Syst. Article, 2015. **5**(19).
10. Rojas, E., et al., *Process mining in healthcare: A literature review.* Journal of Biomedical Informatics, 2016. **61**(April): p. 224-236.
11. Fernandez-Llatas, C., et al., *Process Mining Methodology for Health Process Tracking Using Real-Time Indoor Location Systems.* Sensors, 2015. **15**(12): p. 29821-29840.
12. van der Aalst, W.M.P., *Process Mining: Data Science in Action*. 2 ed. 2016: Springer-Verlag Berlin Heidelberg. 467.
13. National Academies of Sciences, Engineering and Medicine, *Improving Diagnosis in Health Care*. 2015, Washington, DC: The National Academies Press.
14. Jutel, A., *Sociology of diagnosis: a preliminary review.* Sociology of Health & Illness, 2009. **31**(2): p. 278-299.
15. World Health Organization, *Classification of Diseases*. 2019.
16. Johnson, A.E.W., et al., *MIMIC-III, a freely accessible critical care database.* Scientific Data, 2016. **3**: p. 160035-160035.
17. van Eck, M.L., et al., *PM2: A process mining project methodology*, J. Zdravkovic, M. Kirikova, and P. Johannesson, Editors. 2015, Springer, Cham. p. 297-313.
18. Kurniati, A.P., et al., *The assessment of data quality issues for process mining in healthcare using MIMIC-III , a publicly available e-health record database*. 2017.
19. National Center for Health Statistics, *ICD-9-CM Official Guidelines for Coding and Reporting.* 2011.

20. Tenny S, H.M. *Relative Risk*. 10 July 2020; Available from: https://www.ncbi.nlm.nih.gov/books/NBK430824/.
21. Kang, S.-H. and C.W. Ahn, *Tests for the homogeneity of two binomial proportions in extremely unbalanced 2 x 2 contingency tables.* Statistics in medicine, 2008. **27**(14): p. 2524-2535.
22. Mannhardt, F., *Tools & Software — ProM — Event Log Explorer*. 2018.
23. Mannhardt, F., M. De Leoni, and H.A. Reijers. *Heuristic mining revamped: An interactive, data-Aware, and conformance-Aware miner*. in *BPM 2017*. 2017. CEUR-WS.org.
24. Buijs, J.C.A.M., B.F. Van Dongen, and W.M.P. Van Der Aalst, *On the Role of Fitness, Precision, Generalization and Simplicity in Process Discovery*. 2012.
25. Ardiansyah, A., *Replay a Log on Petri Net for Conformance Analysis-plugin.pdf*. 2012.
26. Adriansyah, A., et al., *Measuring precision of modeled behavior.* Information Systems and e-Business Management, 2015. **13**(1): p. 37-67.
27. Kluyver, T., et al., *Jupyter Notebooks—a publishing format for reproducible computational workflows*. 2016. 87-90.
28. Mans, R.S., et al. *Process Mining in Healthcare: Data Challenges when Answering Frequently Posed Questions*. 2012. Berlin, Heidelberg: Springer.
29. Asaria, P., et al., *Acute myocardial infarction hospital admissions and deaths in England: a national follow-back and follow-forward record-linkage study.* The Lancet Public Health, 2017. **2**(4): p. e191-e201.
30. Hall, M., et al., *Multimorbidity and survival for patients with acute myocardial infarction in England and Wales: Latent class analysis of a nationwide population-based cohort.* PLoS Medicine, 2018. **15**(3).
31. Sakr, Y., et al., *Sepsis in Intensive Care Unit Patients: Worldwide Data From the Intensive Care over Nations Audit.* Open forum infectious diseases, 2018. **5**(12): p. ofy313-ofy313.
32. Mehta, R.L., et al., *Sepsis as a cause and consequence of acute kidney injury: Program to Improve Care in Acute Renal Disease.* Intensive care medicine, 2011. **37**(2): p. 241-248.
33. Peerapornratana, S., et al., *Acute kidney injury from sepsis: current concepts, epidemiology, pathophysiology, prevention and treatment.* Kidney international, 2019. **96**(5): p. 1083-1099.
34. Eden, S.V., et al., *Gender and ethnic differences in subarachnoid hemorrhage.* Neurology, 2008. **71**(10): p. 731-735.

The Need for Interactive Data-Driven Process Simulation in Healthcare: A Case Study

Gerhardus van Hulzen(✉)[1][0000−0001−8962−9515],
Niels Martin[1,2,3][0000−0003−3279−3853], and
Benoît Depaire[1][0000−0003−4735−0609]

[1] Hasselt University, Research group Business Informatics, Martelarenlaan 42, 3500 Hasselt, Belgium {gerard.vanhulzen,niels.martin,benoit.depaire}@uhasselt.be
[2] Research Foundation Flanders (FWO), Egmontstraat 5, 1000 Brussels, Belgium
[3] Vrije Universiteit Brussel, Data Analytics Laboratory, Pleinlaan 2, 1050 Brussels, Belgium

Abstract. In healthcare, more and more process execution information is stored in Hospital Information Systems. This data, in conjunction with data-driven process simulation, can be used, e.g. to support hospital management with Capacity Management decisions. However, real-life event logs in healthcare often suffer from data quality issues, affecting the reliability of simulation results. In this work, we illustrate the effects of disregarding data quality issues on simulation outcomes and the importance of domain knowledge using a case study at the radiology department of a hospital. Current literature on data-driven process simulation acknowledges the need for domain expertise but does not provide a framework for conceptualising the involvement of domain experts. Therefore, we propose a novel conceptual framework which interactively involves experts during data-driven simulation model development.

Keywords: data-driven process simulation · data quality · domain knowledge · interactive modelling · healthcare processes

1 Introduction

Worldwide, healthcare systems are under constant pressure. Increasing population numbers, lifestyle factors, ageing populations, and new technologies are the main drivers for increasing healthcare expenses. Simultaneously, healthcare budgets are under pressure due to national budget deficits and savings [14]. Healthcare managers have to improve their care processes to maintain high-quality care for all patients. One key aspect of ensuring this is efficient Capacity Management (CM), which is used to determine the suitable levels of resources, such as equipment, facilities, and staff size [28].

To support hospital management during CM decisions, *Business Process Simulation (BPS)* can be used to determine suitable resource levels objectively. BPS uses a (computer) model to imitate the process. This allows to evaluate the effect of various process modifications without actually implementing them into, nor

disrupting, the real process [21]. For instance, the effect on throughput rates and patient waiting times of installing an additional X-ray scanner can be simulated to determine suitable equipment levels.

Conducting a simulation study is often time-consuming and builds upon subjective inputs, such as interviews and observations. The emerging field of data-driven process simulation in Process Mining (PM) can overcome some of the limitations of "traditional" simulation model development by using data. Data-driven process simulation refers to the automated discovery of a simulation model from process execution data, i.e. an event log [9]. A key challenge in this field is data quality, given its strong impact on the reliability of the simulation results [31]. Because data quality issues are often encountered in healthcare event logs, it is imperative to assess these issues and correct them if needed. This will require domain knowledge. Current literature on data-driven simulation does not provide a clear framework to involve domain experts in model development.

This paper demonstrates the need for interactive data-driven process simulation in healthcare by assessing the impact of data quality issues on simulation results. To this end, a case study at the radiology department of a hospital is considered. In addition, we propose a novel conceptual framework which structures the integration of domain knowledge in the interactive development of data-driven simulation models.

The remainder of this paper is structured as follows. Section 2 gives an overview of the related work. The context of the case study is presented in Section 3. The experimental design, results, and discussion are presented in Section 4. Section 5 introduces our proposed framework for interactive data-driven process simulation. The paper ends with a conclusion in Section 6.

2 Related Work

This work relates to three key domains: (i) simulation for CM decisions in healthcare, (ii) data-driven process simulation, and (iii) data quality in process mining. The following paragraphs give a brief overview of these domains.

Simulation for Capacity Management Decisions in Healthcare. *Capacity Management decisions* in healthcare are concerned with determining the suitable levels of resources, such as staff size, equipment, and facilities [28]. In literature, simulation has been used to determine the required number of beds in general surgery [30]; the number of nurses, doctors, and buffer beds in an Emergency Department (ED) [7]; and the number of computed tomography (CT) scanners in a radiology department [27]. Within the radiology department, the context of our case study, Vieira et al. [32] gave an overview of Operations Research (OR) techniques – which includes simulation – for optimising resource levels and scheduling. For further reference on CM and the use of simulation in healthcare, the reader is referred to one of the existing review papers [26,28,33].

Data-Driven Process Simulation. *Data-driven process simulation* aims to "discover" BPS models from event logs automatically [9]. While existing PM research can support the discovery of individual BPS model components

[19] – e.g. control-flow discovery, decision mining, or organisational mining – less work has been devoted to integrating all these components into a single, simulation-ready model. Rozinat et al. [24] made a first attempt by discovering Coloured Petri Nets (CPNs) to describe the control flow. In addition, gateway routing logic and resource pools were also included. Later, the authors extended their method with activity execution times and case inter-arrival times [25]. Khodyrev and Popova [16] described a similar approach. However, the resource perspective was not included, assuming no resource constraints [16]. Gawin and Marcinkowski [13] provided support for activity durations, control-flow, resources, gateway routing logic, resource schedules, and inter-arrival times. However, the latter two were not automatically derived from data and had to be defined by domain experts [13]. *ClearPath* [15] provides a methodology for discovering and simulating Care Pathways (CPs). Their approach follows an agile, iterative method which facilitates the interaction between the modeller and domain expert, but the obtained process models still have to be manually recreated in their simulation tool *NETIMIS* [15]. *Simod* was the first tool to automatically integrate all components into a single, simulation-ready model to support BPS [6]. In addition, *Simod* is also capable of measuring the accuracy of the derived model and improve it using hyperparameter optimisation [6].

Data Quality in Process Mining. Real-life event logs tend to suffer from *data quality issues*, especially when they originate from flexible environments with substantial manual recording, such as healthcare [5,23]. These issues include missing events and incorrect timestamps, where the latter is often caused by batched registrations by healthcare staff [18,31]. Given the potential impact of event log quality issues on the reliability of PM outcomes, research attention on this topic is increasing. Research efforts are centred around three key topics. Firstly, several frameworks are developed which define event log quality issues [5,29,31]. For instance, Bose et al. [5] define 27 event logs quality issues and group them in four broad classes (i.e. missing, incorrect, imprecise, and irrelevant data). Secondly, research is performed on data quality assessment, targeting the systematic identification of event log quality issues. In this respect, the R-package *DaQAPO* [20], the log query language *QUELI* [1], and the *CP-DQF* [12] for Electronic Health Records (EHRs) provide tools and frameworks to operationalise data quality assessment. They are based on the event log quality issues defined in Vanbrabant et al. [31], Suriadi et al. [29], and Bose et al. [5], respectively. Thirdly, heuristics have been developed which tackle specific data quality issues, e.g. adding missing events [10], imputing missing case identifiers [3], and handling event ordering issues [11].

3 Background: Capacity Management at the Radiology Department

To illustrate the impact of data quality issues in the context of data-driven simulation, a real-life case study is used. This section introduces the case study.

3.1 General Context

The case study relates to a project at the radiology department of a hospital. Hospital management is preparing plans to build new facilities and is requesting input from each department regarding the required capacity. For the radiology department, this relates to the number of examination rooms – i.e. scanners – and the size of the waiting rooms – i.e. the number of seats – for each examination room. The radiology department wants to approach this Capacity Management problem in a data-driven way.

To support this data-driven analysis, process execution data is obtained from the Radiology Information Systems (RIS). This system supports the entire process flow, of which a simplified representation is shown in Fig. 1. The process starts when a patient arrives at the registration desk, after which (s)he is registered. Afterwards, the patient will wait in the waiting room until (s)he is called into the examination room. A nurse helps the patient onto the scanning table and correctly positions the scanner. Next, the image is created. In case the patient needs an additional scan of the same type, e.g. an X-ray scan of both shoulder and neck, this image can be made without leaving the room. After all required scans have been made, the patient can leave the examination room, and the nurse will post-process the images. If the patient still requires additional scans – of a different kind than the previous (e.g. also a CT scan) – (s)he will go to the waiting room of the other examination room. After all scans have been made, the patient can leave the radiology department and return home. Note that the interpretation of the scans by a radiologist is out of scope as it does not impact the required scanner and waiting room capacity.

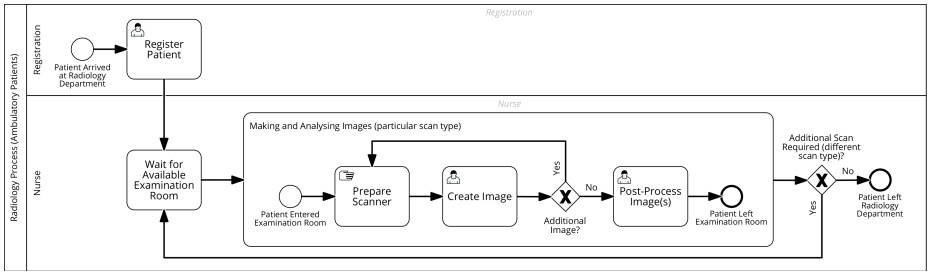


Fig. 1. Simplified process flow of (ambulatory) patients at the radiology department.

To solve the CM problem in this process, *Discrete-Event Simulation (DES)* is used due to the stochastic nature of the process. DES uses simulation to compare policy alternatives before implementing them in practice [33]. Arena v15 [2] was used to simulate the model.

In a DES model, *entities* are dynamic objects which move through the process and trigger the execution of activities [19]. In this case study, entities are patients visiting the radiology department. Four patient types are distinguished:

(i) *ambulatory patients (A)* which are outpatients, (ii) *day hospital patients (D)* which are admitted to the hospital for at most one day, (iii) *hospitalised patients (H)* which are inpatients, and (iv) *emergency patients (S)* which are transferred from the Emergency Department (ED).

The process flow depicted in Fig. 1 actually gives an overview of ambulatory patients. Nevertheless, the flow of the other patient types is, in essence, the same. Only the way patients arrive and where they wait are different. Hospitalised and day hospital patients will wait in their room until they are called in. Emergency patients will wait at the ED.

Depending on the type of scan, a different scanner – and thus a different examination room – is used. In this case study, there are six different types of scans of interest: *angiogram (ANGIO)*, *computed tomography (CT)*, *echocardiogram (ECHO)*, *mammogram (MAMMO)*, *magnetic resonance (MR)*, and *X-ray (RX)*. CT, ECHO, MAMMO, and MR all require separate rooms. ANGIO and RX are performed in RX rooms.

3.2 Data Description

To support the development of the DES model, two years of data from the RIS – from March 2017 until March 2019 – was available. The dataset includes various key timestamps for each patient visit, such as time of registration, and start and end time of scanning. Other attributes, such as the scan type (e.g. ECHO, RX, etc.) and patient type (e.g. ambulatory, emergency, etc.), were also recorded for each patient visit.

The dataset contains 404,750 individual patient visits. The proportions per patient type were 60%, 23%, 15%, and 2% for ambulant, hospitalised, emergency, and day hospital patients, respectively. In total, 464,053 scans were recorded, indicating that the majority of patients only needed one scan. Most scans were RX, i.e. 45%. ECHO represented 19%, followed by MR, 16%, 14% CT, and 5% MAMMO. A very small proportion, less than 0.001%, were ANGIO.

In the process, the activity "Create Image" (cf. Fig. 1) has the most considerable impact on waiting times and throughput rates because it generally takes longer than all other activities. Both start and end timestamps are available of this activity and are recorded when the nurse starts and stops the scanning device, respectively. We initially expected that this activity would not suffer much from quality issues because it is recorded automatically. However, this appeared not to be the case.

Table 1 gives an overview of the scan duration times per scan type. According to the data, some scans took over several years to complete. A few observations even had a negative duration, caused by the end timestamp being recorded before the start timestamp. Given its impact on capacity requirements, the scenario analysis will focus on the effect of scanning time data with data quality issues on simulation outcomes.

Table 1. Scan execution times (in mins).

Scan Type	Min	Max	Mean	Median	SD	IQR
ANGIO	0.00	323,258	14,372.11	26.05	57,336.29	87.83
CT	-726.53	30,605	6.73	1.97	196.86	2.23
ECHO	-79.00	116,685	71.36	23.38	636.37	28.48
MAMMO	-6.48	40,780	16.41	2.98	531.00	1.35
MR	0.00	946,449	161.22	11.48	9,679.90	6.90
RX	-1,031.63	2,109,457	22.69	0.55	5,111.25	1.20

4 Scenario Analysis: The Impact of Data Quality Issues

4.1 Experimental design

To illustrate the impact of data quality issues w.r.t scanning times, we consider two scenarios:

- **Scenario 1 – Direct sampling:** In this scenario, actual observed data is sampled. This is useful when no theoretical distribution, such as the Gaussian, exponential, or gamma distribution, fits the data well. However, the disadvantage is that only the observed values can be used, which is problematic for smaller datasets [17].
- **Scenario 2 – Distribution fitting:** In this scenario, a distribution is fitted to the observed data. We used the distribution with the *least worst* fit because not a single distribution fitted the data well. With this approach, we follow the state-of-the-art of data-driven BPS techniques.

For each scenario, three alternative data filtering approaches are compared:

- **Alternative 1 – Validated filtering (VF):** In this alternative, which is the baseline, we used filtered data validated by domain experts. For scenario 2, we had to use empirical distributions for this alternative as none of the theoretical distribution provided a good fit. In the other two alternatives, we always used theoretical distributions.
- **Alternative 2 – No filtering (NF):** Here, we used the unfiltered data directly. Only observations less than zero were filtered out because the simulation model cannot handle negative activity durations.
- **Alternative 3 – Context-agnostic filtering (CAF):** Even without any domain knowledge, one would immediately notice that the maximum values in Table 1 are unrealistic. Therefore, this alternative uses filtered data to exclude anomalies. We adopted the commonly used *box plot rule* to detect anomalies in the absence of domain knowledge. Any observation smaller than $Q_1 - 1.5IQR$ or larger than $Q_3 + 1.5IQR$ is removed [8]. If the lower limit was less than zero, zero was used instead.

The length of the simulation run was set at two years for each alternative in each scenario. Initial experimentation showed that outliers in Alternative 2 caused severe queue accumulation, which resulted in i.a. extreme waiting times.

Therefore, we integrated a weekly "reset", which removed all patients from queues and ongoing scans. We will refer to this reset as "flushing" and kept track of the weekly number of flushed patients.

To compare the alternatives, we focused on patient throughput and waiting times. Moreover, we looked at the flush count mentioned above. To measure the true effect of the different distributions used in each alternative, *common random number streams (CRNs)* are used. Consequently, the same random numbers are sampled across all alternatives. To compare the difference between alternatives, we used the non-parametric *Wilcoxon-Mann-Whitney (WMW)* test. Instead of using the original observations, ranks are used to compare the difference between two samples. This has the advantage that no underlying distribution is assumed [22]. To control the *false discovery rate (FDR)* of the multiple testing problem, we used the *Benjamini–Yekutieli* procedure [4] to adjust the p-values.

4.2 Results

Throughput Times The *throughput time* measures the elapsed time between the patient's arrival and departure. Because a patient could require multiple scans, the *average throughput time per examination* is considered by dividing the throughput time of a patient by the number of scans. Patients who were "flushed" did not complete all scans and are therefore excluded from this measure.

As shown in Table 2, the throughput times for NF are much higher than VF, e.g. in Scenario 2, the average throughput time per examination for hospitalised patients is almost 100 times longer. The differences between CAF and VF are also statistically significant, albeit much smaller. For day hospital patients, representing 0.5% of the observations for this measure, the differences between VF and CAF were not statistically significant. Nevertheless, important differences in mean throughput times are observed due to larger outlier values for CAF.

Waiting Times The *waiting time* is the time a patient spends in a queue before undergoing a scan. Table 3 shows comparable differences as the throughput times. Again, large differences between VF and NF are observed, e.g. the average waiting time for hospitalised patients is more than 150 times longer in NF than VF for Scenario 2. For day hospital patients, only the difference between VF and CAF in Scenario 1 is not significant, even though the absolute difference between the means is, again, rather large, indicating the presence of outliers.

Flush Counts The more patients are flushed at the end of a week, the more this indicates that queues have accumulated throughout that week. Especially in the NF alternative, many patients have to be flushed to "reset" the process at the end of a week, in some cases even more than a thousand patients in total. The differences between VF and CAF are much smaller, i.e. on average less than one patient more was flushed in CAF. However, it should be noted that sometimes the maximum number of flushed patients in CAF was much higher than in VF, e.g. for Scenario 2, VF flushed at most two hospitalised patients, whereas in CAF this was at most 46. For ambulatory patients, this was smaller, i.e. nine and seventeen, respectively.

Table 2. Throughput times per examination (in min) per patient type (A, D, H, S) and alternative (VF, NF, CAF).

ADHS	Model 1	Model 2	Mean Model 1	Mean Model 2	Adj. p-value	Significance
Scenario 1						
A	VF	NF	25.8493	556.622	<0.0001	****
A	VF	CAF	25.8493	45.8470	<0.0001	****
D	VF	NF	25.0905	129.1714	0.0077	**
D	VF	CAF	25.0905	42.2703	0.6525	ns
H	VF	NF	29.6365	606.8702	<0.0001	****
H	VF	CAF	29.6365	34.1812	<0.0001	****
S	VF	NF	13.4890	92.7292	<0.0001	****
S	VF	CAF	13.4890	16.3655	<0.0001	****
Scenario 2						
A	VF	NF	25.8622	938.1933	<0.0001	****
A	VF	CAF	25.8622	48.8646	<0.0001	****
D	VF	NF	24.6727	363.3950	<0.0001	****
D	VF	CAF	24.6727	118.0489	0.1479	ns
H	VF	NF	29.6764	2,740.0265	<0.0001	****
H	VF	CAF	29.6764	253.3411	<0.0001	****
S	VF	NF	13.5586	76.9503	<0.0001	****
S	VF	CAF	13.5586	17.5182	<0.0001	****

****: p-value < 0.0001, ***: p-value < 0.001, **: p-value < 0.01, *: p-value < 0.05, ns: not signif.

Table 3. Waiting times (in min) per patient type (A, D, H, S) and alternative (VF, NF, CAF).

ADHS	Model 1	Model 2	Mean Model 1	Mean Model 2	Adj. p-value	Significance
Scenario 1						
A	VF	NF	8.7470	536.3071	<0.0001	****
A	VF	CAF	8.7470	26.4227	<0.0001	****
D	VF	NF	10.8402	112.1685	<0.0001	****
D	VF	CAF	10.8402	27.4978	1.0000	ns
H	VF	NF	17.3111	592.3354	<0.0001	****
H	VF	CAF	17.3111	20.7817	<0.0001	****
S	VF	NF	1.7663	76.4688	<0.0001	****
S	VF	CAF	1.7663	3.6703	<0.0001	****
Scenario 2						
A	VF	NF	8.7252	912.6071	<0.0001	****
A	VF	CAF	8.7252	31.3267	<0.0001	****
D	VF	NF	10.3187	337.0036	<0.0001	****
D	VF	CAF	10.3187	99.9275	0.0140	*
H	VF	NF	17.2705	2,737.3727	<0.0001	****
H	VF	CAF	17.2705	243.9300	0.0140	*
S	VF	NF	1.7545	47.0165	<0.0001	****
S	VF	CAF	1.7545	5.3972	<0.0001	****

****: p-value < 0.0001, ***: p-value < 0.001, **: p-value < 0.01, *: p-value < 0.05, ns: not signif.

4.3 Discussion

The results illustrate the need to consider data quality issues seriously. The unfiltered alternative – which completely neglects these issues – exhibits much higher throughput times, waiting times, and flush counts than the validated baseline. The difference between context-agnostic and validated filtering is smaller but still highly relevant. For instance, waiting times for hospitalised patients are up to eight times longer in CAF. However, for other performance metrics, such as flush counts, the differences between VF and CAF are smaller.

In this case study, the cut-off points for outliers in VF and CAF happened to be reasonably close to each other, except for echocardiograms. The domain experts indicated a maximum of 30 mins, whereas the box plot rule returned 84.64 mins. However, this does not give any guarantee for other cases as context-agnostic filtering does not take into account the specificities of a particular domain in any way. Therefore, domain knowledge is always required to achieve accurate simulation results.

When comparing the differences between the two scenarios for each alternative (i.e. comparing the outcomes under direct sampling with their counterpart under distribution fitting), large differences are often observed between throughput and waiting times, even though the same input data was used. A possible explanation is that the theoretical distributions did not fit the data well. Therefore, we highlight the need to report goodness-of-fit (GoF) statistics in state-of-the-art data-driven BPS discovery algorithms and use direct sampling or empirical distributions in case no theoretical distribution fits the data well.

5 Interactive Data-Driven Process Simulation

As illustrated in the case study, data quality issues can have a profound impact on the reliability of simulation results. Moreover, domain knowledge plays a vital role in the development of a simulation model. Without domain knowledge, it is, e.g. challenging to determine whether particular observations are exceptional – but plausible – or data errors. Even though current literature on data-driven process simulation acknowledges the need for domain expertise for i.a. validation purposes, no framework conceptualises how this knowledge should be incorporated.

To enhance the integration of domain knowledge in the development of data-driven simulation models, we propose a novel conceptual framework which interactively involves experts during model building. This framework, which is visualised in Fig. 2, distinguishes three interaction cycles. In the *first cycle*, the initial model is constructed. For each required modelling task (e.g. entity arrival rate, activity durations, resource roles, etc.) – of which an overview is presented in Martin et al. [19] – the data requirements are verified. For instance, mining resource roles requires the presence of a resource attribute. If these requirements are not fulfilled, the domain expert is asked for additional input to perform this modelling task. Conversely, if the requirements are fulfilled, the quality of the

data is assessed, and an applicable discovery algorithm is employed. Next, the results of the discovery algorithm and detected data quality issues are presented for a check by the domain expert. (S)he can then solve any data quality-related issues and tweak the discovery parameters until the results are satisfactory.

The *second cycle* integrates all discovered model components from the first cycle into a single, simulation-ready model. The entire model is simulated, and the domain expert checks the preliminary results. If the simulation outputs do not satisfactorily reflect reality, the model can be "calibrated" by altering the simulation parameters. An estimation of the impact of the altered parameter on simulation outcomes is delivered in real-time, so the expert does not have to wait until the entire simulation has been completed before receiving an indication whether the altered parameter results in the desired change.

The final and *third cycle* is concerned with the validation of the model. The calibrated model from the second cycle is simulated comprehensively and validated by the domain expert. In addition, a validation dataset – which was not used to discover the model – can be used as well. If the desired accuracy level is not achieved, the domain expert can modify the simulation parameters again. The final validated model can be used for the evaluation of various scenarios and further analyses.

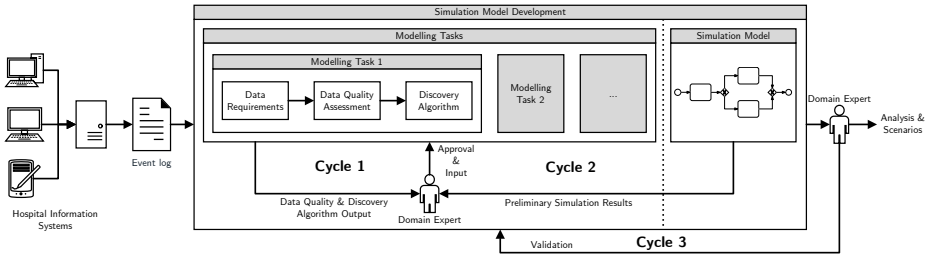


Fig. 2. Interactive data-driven process simulation framework.

6 Conclusion

Data-driven process simulation has great potential within a healthcare context, e.g. to support hospital management with Capacity Management decisions. However, real-life data extracted from Hospital Information Systems tend to suffer from data quality issues, which affects the reliability of simulation results. The presented case study at the radiology department of a hospital illustrates the impact of these issues, as well as the importance of domain knowledge. Current literature on data-driven process simulation acknowledges the need for domain expertise but does not provide a framework to conceptualise the involvement of domain experts. Therefore, we propose a novel conceptual framework which interactively involves experts during data-driven simulation model building. In this

framework, a distinction is made between three cycles: an initial development cycle, a calibration cycle, and a validation cycle.

Future work will focus on how the interaction between the domain expert and the framework will occur more specifically. Ultimately, our goal is to implement our framework into a tool to support the integration of domain knowledge into the development of data-driven process simulation models. In addition, this case study highlights the need for further research on identifying and remedying data quality issues in a healthcare context.

References

1. Andrews, R., Suriadi, S., Ouyang, C., Poppe, E.: Towards Event Log Querying for Data Quality. In: On the Move to Meaningful Internet Systems. LNCS, vol. 11229, pp. 116–134 (2018)
2. Arena. Rockwell Automation, Inc. (2016), https://www.arenasimulation.com/
3. Bayomie, D., Awad, A., Ezat, E.: Correlating Unlabeled Events from Cyclic Business Processes Execution. In: Proceedings of the 28th International Conference on Advanced Information Systems Engineering. LNCS, vol. 9694, pp. 274–289 (2016)
4. Benjamini, Y., Yekutieli, D.: The Control of the False Discovery Rate in Multiple Testing Under Dependency. Ann. Statist. **29**(4), 1165–1188 (2001)
5. Bose, R.P.J.C., Mans, R.S., van der Aalst, W.M.P.: Wanna Improve Process Mining Results? It's High Time We Consider Data Quality Issues Seriously. In: Proceedings of the 2013 IEEE Symposium on Computational Intelligence and Data Mining. pp. 127–134 (2013)
6. Camargo, M., Dumas, M., González-Rojas, O.: Automated Discovery of Business Process Simulation Models from Event Logs. Decis. Support Syst. **134** (2020)
7. Carmen, R., Defraeye, M., Van Nieuwenhuyse, I.: A Decision Support System for Capacity Planning in Emergency Departments. Int. J. Simul. Model **14**(2), 299–312 (2015)
8. Chandola, V., Banerjee, A., Kumar, V.: Anomaly Detection: A Survey. ACM Comput. Surv. **41**(3), 15:1–15:58 (2009)
9. Depaire, B., Martin, N.: Data-Driven Process Simulation. Encyclopedia of Big Data Technologies (2018)
10. Di Francescomarino, C., Ghidini, C., Tessaris, S., Sandoval, I.V.: Completing Workflow Traces Using Action Languages. In: Proceedings of the 27th International Conference on Advanced Information Systems Engineering. LNCS, vol. 9097, pp. 314–330 (2015)
11. Dixit, P.M., Suriadi, S., Andrews, R., Wynn, M.T., ter Hofstede, A.H.M., Buijs, J.C.A.M., van der Aalst, W.M.P.: Detection and Interactive Repair of Event Ordering Imperfection in Process Logs. In: Proceedings of the 30th International Conference on Advanced Information Systems Engineering. LNCS, vol. 10816, pp. 274–290 (2018)
12. Fox, F., Aggarwal, V.R., Whelton, H., Johnson, O.A.: A Data Quality Framework for Process Mining of Electronic Health Record Data. In: Proceedings of the 2018 IEEE International Conference on Healthcare Informatics. pp. 12–21 (2018)
13. Gawin, B., Marcinkowski, B.: How Close to Reality is the "as-is" Business Process Simulation Model? Organizacija **48**(3), 155–175 (2015)
14. Hicks, C., McGovern, T., Prior, G., Smith, I.: Applying Lean Principles to the Design of Healthcare Facilities. Int. J. Prod. Econ. **170**, 677–686 (2015)

15. Johnson, O.A., Ba Dhafari, T., Kurniati, A., Fox, F., Rojas, E.: The ClearPath Method for Care Pathway Process Mining and Simulation. In: Proceedings of the BPM 2018 International Workshops. LNBIP, vol. 342, pp. 239–250 (2019)
16. Khodyrev, I., Popova, S.: Discrete Modeling and Simulation of Business Processes Using Event Logs. In: Proceedings of the 14th International Conference on Computational Science. Procedia Computer Science, vol. 29, pp. 322–331 (2014)
17. Law, A.M.: Simulation Modeling and Analysis. McGraw-Hill Education, New York, NY, USA, fifth edn. (2014)
18. Mans, R.S., van der Aalst, W.M.P., Vanwersch, R.J.B.: Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes. SpringerBriefs in BPM, Springer International Publishing, Cham, Switzerland (2015)
19. Martin, N., Depaire, B., Caris, A.: The Use of Process Mining in Business Process Simulation Model Construction. Bus. Inf. Syst. Eng. **58**(1), 73–87 (2016)
20. Martin, N., Van Houdt, G., Janssenswillen, G.: Towards More Structured Data Quality Assessment in the Process Mining Field: The DaQAPO Package. In: Proceedings of the European R Users Meeting (2020)
21. Melão, N., Pidd, M.: Use of Business Process Simulation: A Survey of Practitioners. J. Oper. Res. Soc. **54**(1), 2–10 (2003)
22. Neuhäuser, M.: Wilcoxon–Mann–Whitney Test. In: International Encyclopedia of Statistical Science, pp. 1656–1658. Springer-Verlag, Berlin, Germany (2011)
23. Rebuge, Á., Ferreira, D.M.R.: Business Process Analysis in Healthcare Environments: A Methodology Based on Process Mining. Inf. Syst. **37**(2), 99–116 (2012)
24. Rozinat, A., Mans, R.S., van der Aalst, W.M.P.: Mining CPN Models: Discovering Process Models with Data from Event Logs. In: Proceedings of the 7th Workshop and Tutorial on Practical Use of Coloured Petri Nets and the CPN Tools. DAIMI PB, vol. 579, pp. 57–76 (2006)
25. Rozinat, A., Mans, R.S., Song, M., van der Aalst, W.M.P.: Discovering Simulation Models. Inf. Syst. **34**(3), 305–327 (2009)
26. Salleh, S., Thokala, P., Brennan, A., Hughes, R., Booth, A.: Simulation Modelling in Healthcare: An Umbrella Review of Systematic Literature Reviews. PharmacoEconomics **35**(9), 937–949 (2017)
27. Shakoor, M.: Using Discrete Event Simulation Approach to Reduce Waiting Times in Computed Tomography Radiology Department. Int. J. Ind. Manuf. Eng. **9**(1), 177–181 (2015)
28. Smith-Daniels, V.L., Schweikhart, S.B., Smith-Daniels, D.E.: Capacity Management in Health Care Services: Review and Future Research Directions. Decis. Sci. **19**(4), 889–919 (1988)
29. Suriadi, S., Andrews, R., ter Hofstede, A.H.M., Wynn, M.T.: Event Log Imperfection Patterns for Process Mining: Towards a Systematic Approach to Cleaning Event Logs. Inf. Syst. **64**, 132–150 (2017)
30. VanBerkel, P.T., Blake, J.T.: A Comprehensive Simulation for Wait Time Reduction and Capacity Planning Applied in General Surgery. Health Care Manag. Sci. **10**(4), 373–385 (2007)
31. Vanbrabant, L., Martin, N., Ramaekers, K., Braekers, K.: Quality of Input Data in Emergency Department Simulations: Framework and Assessment Techniques. Simul. Model. Pract. Theory **91**, 83–101 (2019)
32. Vieira, B., Hans, E.W., van Vliet-Vroegindeweij, C., van de Kamer, J., van Harten, W.: Operations Research for Resource Planning and -Use in Radiotherapy: A Literature Review. BMC Med. Inform. Decis. Mak. **16**(1), 149 (2016)
33. Zhang, X.: Application of Discrete Event Simulation in Health Care: A Systematic Review. BMC Health Serv. Res. **18**(1), 687 (2018)

Process mining on the extended event log to analyse the system usage during healthcare processes (Case study: the GP Tab usage during chemotherapy treatments)

Angelina Prima Kurniati[1,*[0000-0002-4747-1067], Geoff Hall[2[0000-0002-8864-5932], David Hogg[3[0000-0002-6125-9564], Owen Johnson [3[0000-0003-3998-541X]

[1] School of Computing, Telkom University, Bandung, Indonesia, 40257
[2] School of Medicine, University of Leeds, UK, LS2 9JT; Leeds Teaching Hospitals NHS Trust, Leeds, UK, LS9 7TF
[3] School of Computing, University of Leeds, UK, LS2 9JT
lncs@springer.com

Abstract. In healthcare, process mining has been used in many case studies to discover and analyse process models of patient treatments. Process mining is generally applied to analyse the event log of patient treatments as extracted from the Electronic Health Record (EHR). In this study, we proposed an approach to combine the event log of patient treatments with the clinical user access log of the hospital information system to analyse system usage during patient treatments. Our case study combined an event log of breast cancer patients receiving chemotherapy treatments in the Leeds Cancer Centre with the user access log in the hospital information system. The event log of patient records during chemotherapy was extracted from the EHR system. The clinical user access log was extracted from the Splunk, a web-based log management system in the hospital. Combining records from those two logs has been useful to provide information on system usage during patient treatment. Our experiment focused on the GPTab, a functionality that allows clinicians during consultations to check on patient records on their GP visits. We applied both statistical and clinical evaluations to ensure that the findings are statistically correct and clinically meaningful. We captured the phenomena of the decreasing number of patients on the subsequent cycles of chemotherapy and when GPTab has been used during the course of chemotherapy. This approach is potentially useful for general cases to analyse system usage during process execution and can be applied to investigate the effects of system changes to process executions.

Keywords: Process Mining, Extended Event Log, Clinical User Access Log, Chemotherapy, Cancer Treatment, EHR.

1 Introduction

As a large group of diseases, cancer is very complex and can affect any part of the body [1]. There are at least 65 recognised types of cancer [2]. Breast cancer is the most common cancer in women affecting about 12% of women in the world [3]. In the UK, breast cancer is one of the four most common cancer types, along with pros-

tate cancer, lung cancer, and colorectal cancer [4]. Breast cancer [5] is diagnosed by physical exam, mammogram, ultrasound, MRI, blood chemistry studies, and biopsy of the affected area of the breast. Surgery is the primary treatment, which may be followed by chemotherapy or radiation therapy, or both [6]. A course of chemotherapy [7] is usually done in six cycles, where each cycle is given 21 days after the previous one. Some patients might not be able to get a cycle of chemotherapy due to some adverse events, including emergency admission and neutropenia.

Process mining is a process-oriented data science approach that uses event logs for discovering and analysing business process models [8]. An event log is a record of timestamped activities generated automatically by the information system. Process mining has been applied in healthcare processes [9] for quality improvement, patient safety, and resource optimisation in healthcare settings [10]. Our literature review of process mining in Oncology [11], the study of cancer, found the limited availability and accessibility of suitable datasets for process mining. Our earlier study explored a publicly available dataset for process mining in healthcare [12], [13]. In this study, we were fortunate in having access to explore the in-house developed PPM EHR system including the database, the software developers of the system, the training team, clinical staff and senior clinicians involved in the process.

Our case study is based on a de-identified extract from the Patient Pathway Manager (PPM) database of the PPM EHR system [14]. The patient dataset has been used in the previous study to define real-life clinical pathways during chemotherapy [15]. This paper presents a worked example to analyse General Practitioner (GP) Tab usage during chemotherapy treatment on breast cancer patients. GPTab is a menu that allows clinicians to access patient records in the GP system. The GPTab presents clinical information (diagnosis, allergies, medications, etc.) recorded in the registered Leeds GPs. Accessing GP Tab during consultations in chemotherapy cycles improves understanding of patient condition and support decision making for patient treatment. We described an approach to enhance a process model through an extension of the event log, by combining patient records with the user access log. This approach is potentially useful in many other cases to enhance process mining approaches with user access log describing real user accessing information systems.

2 Patient Pathways Manager (PPM) EHR System

The PPM EHR system is used in the Leeds Teaching Hospitals NHS Trust (LTHT), the largest provider of specialised services in England that manages six hospitals, including St James's University Hospital (SJUH) [16]. The SJUH hosts the Leeds Cancer Centre, one of Europe's large cancer centres [17]. The PPM system integrates data from multiple systems within the LTHT, including patient admissions, treatments (chemotherapy, surgery, and radiotherapy), pathology, investigations, Multidisciplinary Team (MDT) meetings, consultations, and outpatients.

The PPM database contains clinical information about all patients within the hospital, including cancer patients. We gained access to the PPM database through an IRAS application that allows direct access to a secure SQL database on a virtual machine. The data has been checked, cleaned, and aggregated before approval for access

by the research team. The PPM database consists of clinical data of more than 3 million patients, of which more than 270,000 patients have at least one cancer-related diagnosis. The PPM EHR system is connected to patient records in other service providers, including General Practitioners (GPs), Mental Health, and Community services. Fig. 1 shows a screenshot of GPTab screen in the PPM EHR system.

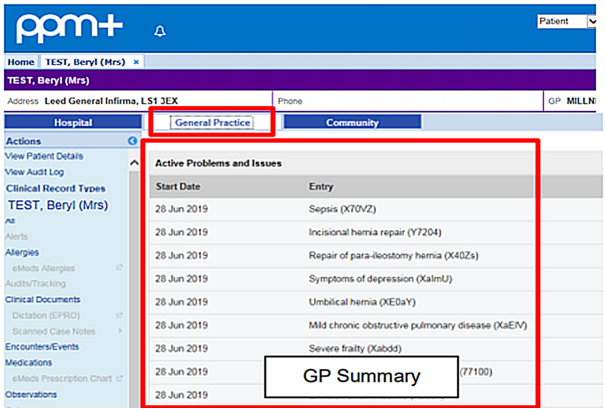


Fig. 1. Screenshot of the GPTab in the PPM EHR system, from the PPM support website [18].

The clinical user access log is recorded in PPM Splunk. The PPM Splunk is web-based application management that captures real-time user access to the PPM system, which is useful in analysing system usage for specific functionalities. Every time a user views data in the PPM EHR system, the system automatically recorded the activity in the PPM Splunk. In this study, the healthcare user access log was focused on the GPTab access log, as a representative of functionalities related to cancer treatment. GPTab is a functionality that can be used by clinicians to access patient records in the GP system, to support clinical decisions related to patient treatment.

3 Methodology

The general methodology is based on the Process Mining Project Methodology (PM^2) [19] with a focus on the *Mining and analysis* step (Fig. 2).

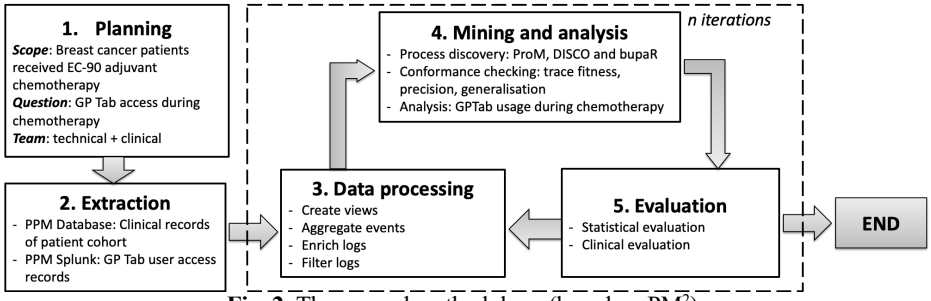


Fig. 2. The general methodology (based on PM^2)

We did the stages in the methodology in at least two iterations: once with only the clinical records as the input, and once with a combination of the clinical records and the healthcare user access log. For simplicity and ease-of-understanding, this paper describes only the final iteration and summarise the findings in the intermediate iterations as part of the final iteration.

The Planning stage identified the scope, the team, and the research questions in the study. The scope of this study was to analyse GPTab usage during chemotherapy treatment of breast cancer patients in the PPM system. The research questions were:

Q1. What are the most followed paths and the exceptional paths?

Q2. How did clinicians use GPTab during the course of chemotherapy?

Our team consisted of process mining experts, clinical experts, representatives of the development and training teams of the PPM EHR system. We did at least one meeting in each stage of the study to discuss the plan, progress of the study, and validation of the findings. The discussion was done to ensure domain expert engagement during all stages of the study, as suggested in the ClearPath method [20].

The Extraction stage included the patient clinical records from the PPM database and the user access log from PPM Splunk. *The patient clinical records* are included if (1) the patient had at least one diagnosis of breast cancer (ICD-10 C50) and received epirubicin and cyclophosphamide (EC90) chemotherapy as adjuvant treatment and (2) the patient was first diagnosed with breast cancer between 2014 and 2018. The EC90 is one of the most commonly used regimens in Leeds Cancer Centre in the specified time period. *The GPTab user access records* from PPM Splunk are included if clinicians access GP records of patients in the cohort during their cancer treatment between 2014 (when GPTab was introduced) and 2018. Combining patient clinical records with user access records is useful to get additional data from user access log that is not recorded in the patient clinical records, in this study, adding GPTab access activity to the chemotherapy pathways. The extraction stage is illustrated in Fig. 3.

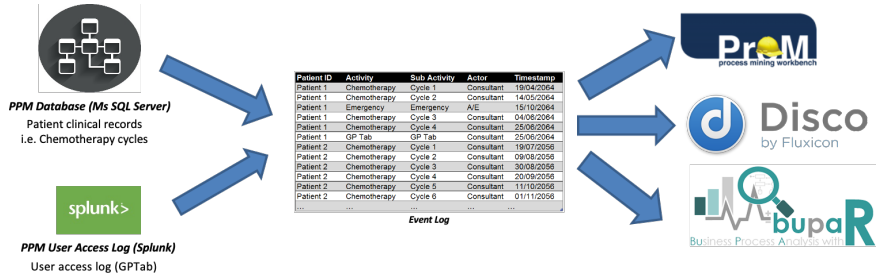


Fig. 3. The extraction stage, combining patient clinical records with user access records.

The Data Processing stage consisted of creating views, aggregating events, enriching logs, and filtering logs. The views were created by focusing on the chemotherapy cycles of breast cancer patients. Instead of aggregating events, we used the fine-grained event names, which are Cycle 1, Cycle 2, up to Cycle 6, representing the cycle number of chemotherapy. Log enrichment added information to the event log, in this case, the process duration for each patient that was calculated as the number of days from the first activity to the last one in the recorded treatment. We also included Emergency and Neutropenia events as suggested by clinical experts to be the two

events potentially affecting chemotherapy progressions. We extracted the Emergency events as they were recorded in the Admission table with an Emergency Admission type. Neutropenia is a condition where a patient had a neutrophil count less than normal (<1.5 x 10^9/L). More details about those two additional events had been described in our previous study using the same dataset [15]. An attribute-based log filtering was done by filtering in selected events to include only the chemotherapy cycle events of patient treatment. The patient records were transformed into an event log, which contains {case_id, activity, resource, timestamp}. The event log was loaded into ProM tools and R for analysis in the next stage.

The Mining and Analysis stage included process discovery, conformance checking [22], enhancement, and process analytics. Process discovery was done in the fine-grained level to model chemotherapy cycles of patients in the selected cohort. The adjuvant chemotherapy for breast cancer patients is commonly given in six cycles, sequentially from Cycle 1 to Cycle 6. The main tools for process discovery were ProM 6.8 [24], DISCO [25], and bupaR [26]. ProM is an academic platform that is widely used in process mining projects. DISCO is used in this study to get an early model easily, based on the fuzzy miner algorithm. bupaR is a library in R that was used in this study to support a more detailed statistical analysis.

Enhancement was done by extending the event log of the patient records with the GPTab access log in the PPM Splunk. Fig. 4 shows a screenshot containing detailed data on the date and time, page address, patient id and user id recording a time when a clinician had accessed the GP Tab page of a patient. There is also a bar chart visualising the number of records on a daily basis. The bar chart shows an obvious pattern of weekday- and weekend- usages.

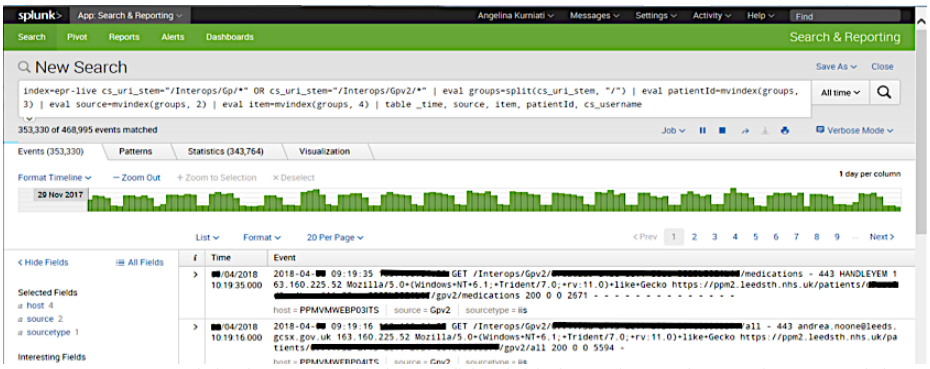


Fig. 4. A query result in the PPM Splunk. Confidential information such as Patient ID and dates are blocked in black.

The Evaluation stage was done to diagnose, verify, and validate the results of the previous stages. In this study, the evaluation analysed all findings from the statistics and clinical perspectives. The statistical evaluation was done to verify and validate the result quantitatively, which was later confirmed to the clinical experts and the representative of the development team. The clinical evaluation was done to make sure that the findings reflected reality, supported and enhanced prior knowledge of the clinical experts about patient treatment.

4 Results and Discussion

4.1 The Extracted Data

We extracted Leeds patients diagnosed with breast cancer (C50) who received EC-90 as adjuvant chemotherapy, whose GP Tab was accessed by clinicians from 2014 to 2018. There were 738 patients included in this selection. Table 1 shows a list of the eight selected events for process discovery, which consists of six cycles of chemotherapy and two adverse events (emergency admission and neutropenia).

Table 1. Selected Events for Process Discovery

Event	Cycle						Emergency	Neutropenia
name	1	2	3	4	5	6		
Patients (n)	738	725	699	487	402	380	380	412
Percentage	100%	99%	95%	66%	55%	52%	-	-
Med (days)	21	21	21	21	21	-	-	-

Table 1 shows that 738 patients received Cycle 1 of chemotherapy, but the number decreases in the following cycles. The median duration from a Cycle to the next one is 21 days, which reflects the typical duration of treatment in reality. This finding has been discussed with clinical experts. It has been confirmed to reflect the reality where patients might find several conditions that prevent them from completing the course of chemotherapy. It is shown that among patients who started receiving *Cycle 1* of EC-90 as adjuvant chemotherapy, only around half of them (n=380; 52%) completed *Cycle 6*. This condition needs to be explored more, to learn what were the possible conditions preventing patients from completing the treatment.

4.2 Discovered Process Models and the Conformance

We presented Table 2 to show the 15 most common trace variants out of 289 variants in total. Each of those 15 variants followed by at least seven patients.

Table 2. Top Eight Trace Variants

Var	Trace Variant	n	(%)
1	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4 - Cycle 5 - Cycle 6*	120	16.26
2	*Cycle 1 - Cycle 2 - Cycle 3*	56	7.59
3	*Cycle 1 - Cycle 2 - Cycle 3 – Emergency*	37	5.01
4	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4 - Cycle 5 - Cycle 6 - Emergency*	25	3.39
5	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4*	14	1.90
6	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4 - Cycle 5 - Neutropenic - Cycle 6*	11	1.49
7	*Cycle 1 - Neutropenic - Cycle 2 - Neutropenic - Cycle 3 - Neutropenic*	10	1.36
8	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4 - Neutropenic - Cycle 5 - Cycle 6*	10	1.36
9	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4 - Emergency*	9	1.22
10	*Cycle 1 - Cycle 2 - Cycle 3 – Emergency - Neutropenic*	9	1.22
11	*Cycle 1 - Cycle 2 - Cycle 3 – Neutropenic*	8	1.08
12	*Cycle 1 - Cycle 2 - Cycle 3 - Cycle 4 - Cycle 5*	8	1.08
13	*Cycle 1 - Cycle 2 -Cycle 3 -Neutropenic -Cycle 4 -Cycle 5 -Cycle 6 -Emergency*	8	1.08
14	*Cycle 1 - Cycle 2 - Cycle 3 - Neutropenic - Emergency*	8	1.08
15	*Cycle 1 - Cycle 2 - Cycle 3 - Neutropenic - Cycle 4 - Cycle 5 - Cycle 6*	7	0.95

Table 2 shows that the most common variant is a sequence of *Cycle 1* to *Cycle* 6 (n=120; 16.26%), followed by the second variant that is a sequence of *Cycle 1* to *Cycle 3* (n=56; 7.59%). Our clinical experts confirmed that even though a complete sequence of *Cycle 1* to *Cycle 6* is expected, a lot of patients needed a consultation after *Cycle 3* to decide if the chemotherapy regimen can be continued. Patients might also change regimen after *Cycle 3* and therefore are not captured in this study.

Fig. 5 shows a dotted chart of routine chemotherapy cycles of patients treatments of up to 7 years. The chart shows groups of patients who had not completed six cycles of chemotherapy (the one-third top part of the chart), who completed six cycles of chemotherapy (the middle part), and who had more complicated courses of treatment (the bottom part). In total, 51% (n=376) patients completed all six cycles, without any acute event (n=158; 21%) or having at least one acute event including emergency admission or neutropenia (n=218; 30%). The patients who did not complete six cycles (n=392; 49%), might had acute events (n=207; 28%) or not completing for other reasons (n=155; 21%). Based on our discussion with clinical experts, some of those reasons are missing appointments, disease complications, and personal reasons.

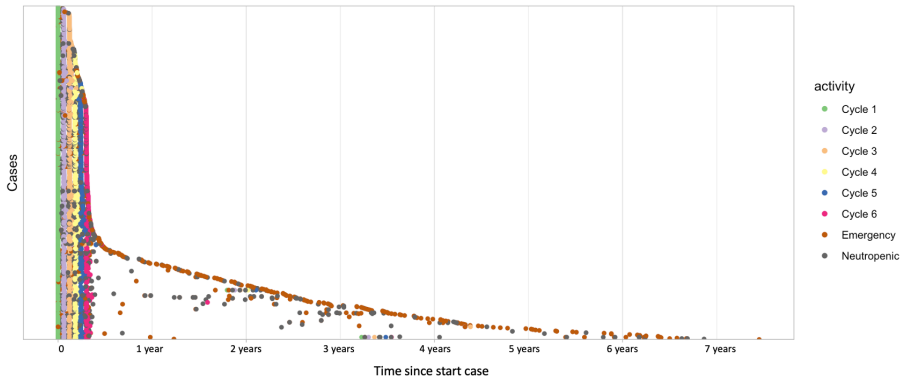


Fig. 5. Dotted chart showing adverse events during six chemotherapy cycles. The x-axis shows duration from the first activity to the last one. The y-axis shows patient id, sorted by durations.

This dotted chart had been shown to the clinicians and all of them agreed that this visualisation helped them understanding the situation more clearly. There are only about a third of patients had the normal and 'happy' path of six cycles of chemotherapy, while the others had incomplete or overly complicated paths of treatment. Some example of patients were picked and discussed with clinical experts to see specific cases where patient conditions preventing them from completing the treatment. Those specific cases are not presented in this paper because presenting data of a small number of patient would breach ethical approvals.

Further analysis of the result was examining the cycles leading to an emergency admission or a neutropenic condition. Table 3 shows that most patients who had emergency admission got it after *Cycle 3* (n=117; 16%), *Cycle 6* (n=90; 12%), or *Cycle 1* (n=81; 11%); while most patients who had *Neutropenic* got it after *Cycle 3* (n=142; 19%), *Cycle 2* (n=123; 17%), or *Cycle 1* (n=94; 13%). Collectively, adverse events (Emergency or Neutropenic) have mostly occurred after *Cycle 3*. Table 2

summarised the pattern of the cycles leading to an acute event and might have a one-to-many relation to trace variants presented in Table 2. For example, *Cycle 3* leading to a *Neutropenic* event in Table 3 (n=142; 19%) is related to variants 7, 11, 13, 14, 15 and other infrequent variants in Table 2.

Table 3. The Cycles Leading to an Acute Event

Activity	leads to Emergency		leads to Neutropenic	
	N (%)	med; mean	N (%)	med; mean
Cycle 1	81 (11)	8 d; 18.4 d	94 (13)	19 d; 23.1 d
Cycle 2	52 (7)	8 d; 43.9 d	123 (17)	19 d; 20.6 d
Cycle 3	**117** (16)	28 d; 27.3 w	**142** (19)	18 d; 61.1 d
Cycle 4	64 (9)	14d; 27.3 w	84 (11)	19 d; 16 d
Cycle 5	22 (3)	13.5 d; 19.2 w	70 (9)	19 d; 33.5 d
Cycle 6	-	-	-	-

It is also important to note that the median and mean duration of acute events after a chemotherapy cycle are generally under 21 days, within the expected duration of a cycle to the next one. This means that patients experienced one or more acute events before the next cycle of chemotherapy, got treated, and continue to the next cycle of chemotherapy as planned. On the last row, Emergency and Neutropenic events after Cycle 6 are not presented because they are not part of this study.

4.3 The Enhanced Process Model

There were 339 out of 738 patients (46%) who had their GPTab accessed by clinicians. This percentage is higher than the percentage of all cancer patients who had their GPTab accessed by clinicians (46,547 out of 339,127 patients; 37%), which showed that clinicians made use of the patient records in the GPTab to support their decisions on the next treatment for their patients. Fig. 6 shows the process model containing the flow from Cycle 1 to Cycle 6 of chemotherapy. During the course of chemotherapy, the GPTab might be accessed by clinicians. The most frequent sequence is that GPTab was accessed after Cycle 6 (n=160; 47%), followed by GPTab access after Cycle 3 (n=110, 32%) and GPTab access after Cycle 4 (n=31; 9%).

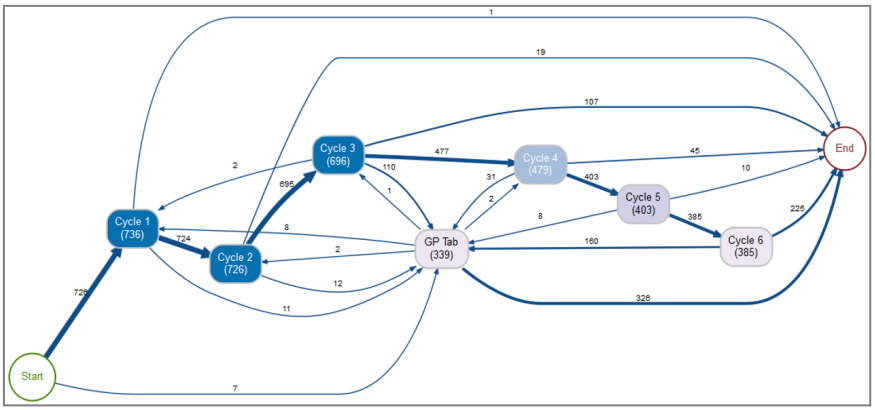


Fig. 6. Process model showing GPTab access during chemotherapy cycles (bupaR). It shows that GPTab was mostly accessed after *Cycle 3*, *Cycle 6*, or *Cycle 4.*

These results have been confirmed by the clinical experts to reflect reality. The clinicians are most likely need to check on patient records in GPTab after the sixth cycle to decide whether to discharge the patient, to follow on the next cycle of chemotherapy, or to suggest another treatment. Clinicians might need to check on patient records in GPTab after Cycle 3, to decide if the next cycles should be delivered as planned or not. Another finding was that GPTab click is mostly the last activity in the pathways, or at the end of treatment (n=326; 96%). The enhanced process model revealed some important insights into how GPTab has been used during the treatment process.

4.4 Process Analytics

Process analytics was done to analyse GPTab usage chemotherapy. This was based on a discussion with a representative of the PPM development team who mentioned that the GPTab had been through some changes during the study period. We followed up this discussion by exploring the increasing pattern of GPTab usage over time. Fig. 7 shows a bar chart of the number of GPTab clicks from July 2014 to December 2018.

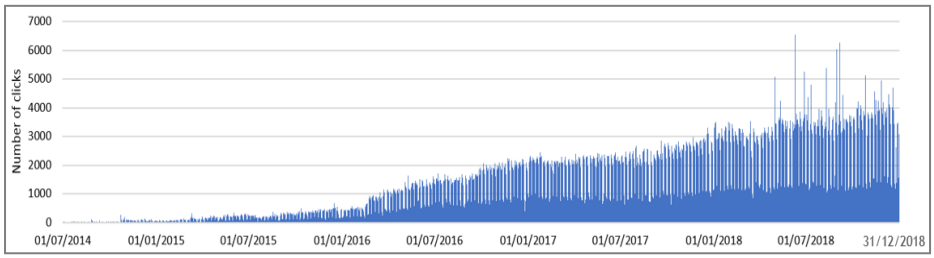


Fig. 7. GPTab clicks each day. It shows that the number of clicks generally increased over time, with steady fluctuations showing the pattern of weekday- and weekend- usages.

Further exploration of the PPM Splunk records shows that in March 2018, the first version of GPTab (GPv1) has been replaced by the second version (GPv2). In September 2017 to February 2018 both versions were accessed by clinicians, and this has been confirmed as the transition period. The transition period from GPv1 to GPv2 can be captured in the monthly usage from 2017 to 2018, as shown in Fig. 8. This has not been seen in Fig. 7, which shows that the transition from the first version to the second one has been done smoothly.

4.5 Statistical and clinical evaluation

The evaluation was done in both statistical and clinical aspects. Statistical evaluation was done throughout the stages by analysing the occurrence numbers and percentages of events in the process. This has been presented in the relevant steps in the previous sections of this paper.

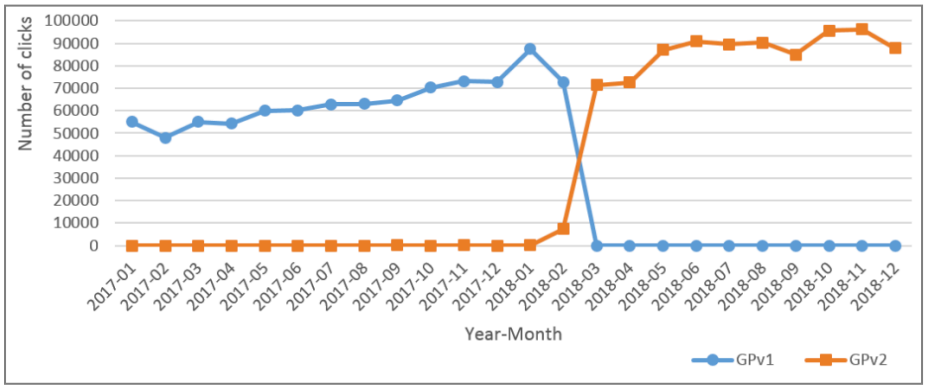


Fig. 8. Monthly usage of GPTab during 2017-2018. The blue dots are monthly usage of the first version (GPv1) and the orange dots are those of the second version (GPv2).

Clinical evaluation was done through discussion with clinical experts. In the Planning stage, clinical experts suggested the scope of the study. The GPTab functionality was chosen based on the availability of the related data to enhance process model of patient treatment. One important insight from the software training team was that for some new features introduced in the PPM software, there was a period when training was given to the clinicians to introduced the use of the new feature, such as GPTab. During the Extraction stage, clinical experts evaluated and suggested details the extraction step. One important suggestion in this stage was the specific type of chemotherapy for breast cancer selected in this study, which is EC90 for adjuvant treatment. In the Data processing stage, clinicians suggested focusing on the effect of the GPTab introduction to the chemotherapy cycles. The findings from the Mining and analysis stage have been discussed with clinical experts. Some of their comments had been presented in the relevant part in Section 4.1 to Section 4.4. The GPTab supported clinicians to decide on the next treatment suitable for their patients, such as to follow with the next cycle of chemotherapy, to change the regimen of chemotherapy, or to discharge the patient.

5 Conclusion

This paper described a process analytics approach by combining patient clinical records with user access log to analyse system usage during patient treatment. A case study presented in this paper was GPTab usage during chemotherapy treatment. Two research questions had been established and answered through a structured experiment following the PM2 stages. The first research question has been answered in the Mining and analysis stage, specifically in the process model (see 4.2). Additional analysis to support this answer has been presented in a trace variant list (Table 2) and a dotted chart (Fig. 5). The second research question has been answered by the enhanced process model (Fig. 6) which shows how GPTab has been used to support clinician to decide the next treatment for their patients. General comments of the findings throughout the stages are that process mining is potentially useful to improve clinical pathway analysis by providing visualisation of process models and additional

results such as trace variance diagrams and dotted charts. Those visualisations supported discussions with the multi-disciplinary team.

Some limitations and potential improvements in this study are as follow. The first is to explore the aggregated events to see how chemotherapy has been given in the sequence from a referral, diagnosis, and a set of treatments. Second, the idea of combining user access records in PPM Splunk with the treatment records in the PPM database was good to analyse the effect of system functionality to the treatment process. Another possibility discussed was to analyse PPM Splunk separately to be compared to the discovered process model from the patient records. Since PPM Splunk recorded all actions done by clinicians during patient treatment, the treatment process itself should be reflected in the records. Third, the extraction and data processing in this study relied on the selection of the best set of events of the specific cohort of patients, based on the understanding of the data and problem domain. Further improvement might be to explore possible ways to select the best set of events based on the data attributes, with less dependence on clinical expert judgments.

Acknowledgment

This research was supported by ClearPath Connected Health Cities Project and was partly funded by the Indonesia Endowment Fund for Education (LPDP). Access to data used in this study is under the Health Research Authority (HRA) Approval Number 206843.

References

[1] American Cancer Society, “The History of Cancer,” 2011. www.cancer.net/patient/Advocacy and Policy/Treatment_Advances_Timeline.pdf (accessed Aug. 09, 2016).

[2] CRUK, “Your cancer type,” *Cancer Research UK*, 2014. http://www.cancerresearchuk.org/about-cancer/type/ (accessed Aug. 09, 2016).

[3] National Cancer Institute, “Breast Cancer - Patient Version,” 2018. https://www.cancer.gov/types/breast/patient/breast-treatment-pdq.

[4] Office of National Statistics, “Cancer registration statistics, England:2016,” 2016.

[5] National Institute for Health and Care Excellence, “Advanced breast cancer overview - NICE Pathways,” *NICE Pathways*, 2016. https://pathways.nice.org.uk/pathways/advanced-breast-cancer.

[6] National Chemotherapy Advisory Group, “Chemotherapy Services in England : Ensuring quality and safety,” *Dep. Heal.*, no. August, 2009.

[7] Royal Cornwall Hospitals NHS Trust, “Clinical Guideline for the Assessment and Management of Chemotherapy Induced Diarrhoea,” pp. 1–11.

[8] W. M. P. van der Aalst, *Process Mining: Data Science in Action*, 2nd ed. Springer-Verlag Berlin Heidelberg, 2016.

[9] E. Rojas and J. Munoz-Gama, “Process mining in healthcare: A literature review,” *J. Biomed. Inform.*, vol. 61, pp. 224–236, 2016.

[10] R. Mans and W. M. P. van der Aalst, “Process mining in healthcare: Data challenges when answering frequently posed questions,” *Process Support Knowl. Represent. Heal. Care*, pp. 140–153, 2013.

[11] A. P. Kurniati, O. Johnson, D. Hogg, and G. Hall, “Process mining in oncology: A literature review,” in *Proceedings of the 6th ICICM* 2016, pp. 291–297, doi: 10.1109/INFOCOMAN.2016.7784260.

[12] A. P. Kurniati, E. Rojas, D. Hogg, and O. Johnson, “The assessment of data quality issues for process mining in healthcare using MIMIC-III , a publicly available e-health record database,” no. 2, 2017.

[13] A. P. Kurniati, G. Hall, D. Hogg, and O. Johnson, “Process Mining in Oncology using the MIMIC-III Dataset,” *IOP J. Phys. Conf. Ser. 971*, vol. 971, no. 012008, pp. 1–10, 2018.

[14] A. Newsham, C. Johnston, and G. Hall, “Development of an advanced database for clinical trials integrated with an electronic patient record system,” *Comput. Biol. Med.*, vol. 41, no. 8, pp. 575–586, 2011.

[15] K. Baker *et al.*, “Process mining routinely collected electronic health records to define real-life clinical pathways during chemotherapy,” *Int. J. Med. Inform.*, vol. 103, pp. 32–41, 2017, doi: 10.1016/j.ijmedinf.2017.03.011.

[16] Leeds Teaching Hospitals NHS Trust, “Leeds Teaching Hospital,” 2016. http://www.leedsth.nhs.uk/ (accessed Jul. 26, 2016).

[17] W. Hazell, “Analysed: The biggest NHS providers of specialised services | News | Health Service Journal,” 2015. https://www.hsj.co.uk/home/analysed-the-biggest-nhs-providers-of-specialised-services/5091147.article (accessed Jul. 30, 2019).

[18] LTHT, “Leeds Care Records GP Tab,” Leeds, 2019. [Online]. Available: https://www.leedscarerecord.org/lcr/widget/whats-in/gp-tab.pdf.

[19] M. L. van Eck, X. Lu, S. J. J. Leemans, and W. M. P. van Der Aalst, “PM2: A process mining project methodology,” in *International Conference on Advanced Information Systems Engineering*, 2015, pp. 297–313, doi: 10.1007/978-3-319-19069-3_19.

[20] O. A. Johnson, T. B. A. Dhafari, A. Kurniati, and E. Rojas, “The ClearPath Method for Care Pathway Process Mining and Simulation,” in *Lecture Notes in Business Information Processing*, 2018, pp. 1–12.

[21] A. Adriansyah, “Replay a Log on Petri Net for Conformance Analysis plug-in.pdf | Algorithms | Logarithm.” Accessed: Oct. 31, 2017. [Online]. Available: https://www.scribd.com/doc/225913652/Replay-a-Log-on-PN-for-Conformance-Analysis-plug-in-pdf.

[22] J. C. A. M. Buijs, B. F. van Dongen, and W. M. P. van Der Aalst, “On the role of fitness, precision, generalization and simplicity in process discovery,” in *OTM Confederated International Conferences*, 2012, pp. 305–322, doi: 10.1007/978-3-642-33606-5_19.

[23] W. M. P. van der Aalst, A. Adriansyah, and B. van Dongen, “Replaying history on process models for conformance checking and performance analysis,” *Wiley Interdiscip. Rev.Data Min.Knowl. Discov.*, vol. 2, no. 2, pp. 182–192, 2012, doi: 10.1002/widm.1045.

[24] W. M. P. van der Aalst, B. F. Van Dongen, C. Gunther, A. Rozinat, H. M. W. Verbeek, and A. J. M. M. Weijters, “ProM: The process mining toolkit,” *CEUR Workshop Proc.*, vol. 489, 2009.

[25] C. W. Günther and A. Rozinat, “Disco: Discover Your Processes.,” in *BPM 2012 Demonstration Track*, 2012, vol. 940, pp. 40–44.

[26] G. Janssenswillen, “bupaR: Business Process Analysis in R,” *R package version 0.4.2*, 2019. https://cran.r-project.org/package=bupaR.

Process Mining on FHIR - An Open Standards-Based Process Analytics Approach for Healthcare

Emmanuel Helm[1,3][0000−0002−1323−3511], Oliver Krauss[1,3][0000−0002−8136−2606], Anna Lin[1][0000−0003−4840−8788], Andreas Pointner[1][0000−0001−8642−1161], Andreas Schuler[2,3][0000−0003−1074−3222], and Josef Küng[3]

[1] Research Department Advanced Information Systems and Technology, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
[2] Department of Medical and Bioinformatics, University of Applied Sciences Upper Austria, 4232 Hagenberg, Austria
firstname.lastname@fh-hagenberg.at
https://aist.fh-hagenberg.at
[3] Johannes Kepler University, 4040 Linz, Austria
josef.kueng@jku.at

Abstract. Process mining has become its own research discipline over the last years, providing ways to analyze business processes based on event logs. In healthcare, the characteristics of organizational and treatment processes, especially regarding heterogeneous data sources, make it hard to apply process mining techniques. This work presents an approach to utilize established standards for accessing the audit trails of healthcare information systems and provides automated mapping to an event log format suitable for process mining. It also presents a way to simulate healthcare processes and uses it to validate the approach.

Keywords: Process Mining · Healthcare · HL7 FHIR.

1 Introduction

We provide a process analytics approach to enable the mining of standardized audit trails of healthcare information systems by transforming them into eXtensible Event Stream (XES) logs via an automated mapping approach. We tested it by simulating a radiology practice workflow, and analyzed the results with a process mining tool.

With diverse use cases and different approaches, techniques, and algorithms, process mining became its own scientific discipline over the last 20 years [1]. With the goal of understanding and improving the real-world processes, process mining provides an evidence-based (i.e., data-driven) view on the processes recorded by information systems. An increasing number of case studies also show the applicability of process mining in the healthcare domain (cf. the reviews in [4, 18]). Most of those case studies focus their analysis on single hospitals or even departments due to problems of data integration or data availability [4].

1.1 Problem Statement

Rebuge and Ferreira [17] conclude in their work that healthcare processes, both organizational and medical treatment, are *highly dynamic*, *highly complex*, *increasingly multi-disciplinary* and *generally ad-hoc*. All four characteristics make it hard to apply process mining techniques. In this work we focus on the aspect of high complexity, partly caused by the high number of participants, heterogeneous information systems, and the resulting lack of interoperability [4, 17].

Rojas et al. [18] found in their review that three implementation strategies for process mining projects in healthcare exist: (1) The majority of case studies work with *direct implementations*, where data is gathered directly from hospital information systems (HIS) for building an event log. Data extraction and building the correct event log poses major challenges here. (2) The second, *semi-automated*, strategy involves the integration and extraction of data from different sources via custom-made developments. The disadvantage here is the ad-hoc, proprietary nature of these developments, as they only work for specific data sources and environments. Both strategies, direct implementation and semi-automated, share the need to understand process mining tools and algorithms for conducting process analytics. (3) The third strategy is the implementation of an *integrated suite*. Specific data sources are connected and integrated, and specific process mining algorithms are executed in order to perform defined analytics tasks. Once implemented, these solutions are easily applicable, but like the semi-automated strategy, fail to integrate other data sources and environments.

We conclude, that a major problem with starting a process mining project in healthcare is that one has to choose between either complex manual data extraction and integration, or locking oneself in on specific data sources and environments (i.e., vendor lock-in).

1.2 Related Work

To overcome the problems of process mining on heterogenous data sources in healthcare, some studies tried to analyze standardized audit trails [3, 7, 16]. We will build on this work, using their concepts of audit events, mapping strategies, and multi-perspective process mining.

Cruz-Correia et al. [3] were the first to explicitly make the connection between standardized auditing in healthcare and process mining. They specifically looked at the Integrating the Healthcare Enterprise (IHE) integration profile Audit Trail and Node Authentication (ATNA). Being one of the core profiles dealing with IT infrastructure in healthcare, ATNA defines how to build up a secure domain that provides patient information confidentiality, data integrity, and user accountability. They analyzed ATNA audit trails from four different hospitals in Portugal and identified several data quality issues.

Later, Helm and Paster [7] investigated the suitability of event logs recorded by the means of IHE ATNA for process mining. They adopted a direct mapping approach, transforming IHE audit messages into XES event logs. They encountered issues regarding the determination of trace identifiers and semantics preserving mapping.

De Murillas et al. [16] took on the previous approach [7] and presented a method to overcome the problems of trace identification and incorrect mappings. By integrating the audit trail data into a generic meta model (OpenSLEX), they provided the means to query and analyze the data from different perspectives.

While these approaches try to solve the issue of heterogeneous data sources, they either lock the user in on a predefined mapping [7] or provide a non-standardized interface to the process data [16] – two shortcomings that can be avoided with our approach.

1.3 Proposed Solution

Supporting definition, instantiation, and execution of workflows is still a topic of vivid discussions in the respective standards development working groups. For the analysis part, first steps have been taken. Standardized Operational Log of Events (SOLE) is a recently developed IHE integration profile. It is a supplement for the radiology technical framework and currently in revision 1.2, published for trial implementation in mid 2018 [13]. SOLE describes the capture and retrieval of operational events in the radiology domain and utilizes transactions from the ATNA profile, including the new RESTful ATNA [12], based on the Health Level Seven (HL7) standard Fast Healthcare Interoperability Resources (FHIR). The profile authors' incentive for writing the SOLE integration profile was the strong desire of healthcare providers "to increase throughput and efficiency, both to improve the quality and timeliness of care and to control costs" [13]. They conclude, that workflow events must be captured in order to be able to apply *business intelligence tools* [13].

We propose an open standards-based process analytics approach for healthcare information systems to overcome the problems mentioned above. It enables the development of tools that combine the easy applicability of an integrated suite with the ability to integrate different data sources. This will make existing process mining tools the *business intelligence tools* the community wants.

To this end, this paper aims to show how existing concepts can be utilized and what changes in the standard are necessary to enable process mining based on HL7 FHIR. This paper also contributes to the field by presenting a novel approach to utilize a process simulation tool in a healthcare environment.

2 Background

This section provides a brief overview on the two major standards involved in building the open process analytics approach, HL7 FHIR and XES.

2.1 HL7 FHIR

FHIR[4] is the latest addition to the family of healthcare interoperability standards maintained and published by HL7 International [8]. FHIR provides a

[4] HL7, FHIR and the FHIR logo are the registered trademarks of Health Level Seven International and their use does not constitute endorsement by HL7.

comprehensive information model which is geared towards supporting semantic interoperability of clinical data. The fundamental building blocks for this information model are *resources*. A resource as described by Mandel et al. [15] is a coherent expression of clinical data and is based on a set of well-defined fields and data types. Every resource comprises the standard defined data content, a human-readable representation of respective content and has an identity. The FHIR specification defines resources for common clinical concepts, e.g., Patient, Medication, Observation, Condition. Besides that, FHIR leverages modern web technologies together with a strong foundation of web standards and offers support for RESTful architectures. Following the RESTful paradigm, FHIR allows to alter the state of a particular resource using a set of predefined actions for Create/Read/Update/Delete (CRUD). If required by a given use-case, it is also possible to apply a more Remote Procedure Call (RPC)-like interaction paradigm. This is achieved by defining operations that work on input and produce an output [9]. The operations can be executed on the server level, on the resource type level, or on the instance level of a specific resource and are typically invoked by a HTTP POST or can alternatively be invoked by a HTTP GET if no changes are caused on the server.

According to HL7 International [8], a central challenge for the FHIR specification is handling the wide variety and variability in diverse healthcare processes. This challenge is solved by offering a simple framework for extending the existing resources and describe use cases based on profiles. Profiling a resource allows to constrain and extend a resource specification for a given context [15]. By providing reference implementations for the specification, HL7 intends to reduce the entry barrier for developing FHIR conformant solutions. The development of the specification and the standard follows a developer first approach, which is reflected by the specification as a mixed standard comprising normative portions and parts still undergoing trial use [8].

2.2 XES

Log data is created from a variety of different systems with their own proprietary data models, formats, and semantics. Process mining techniques require their input data in a specific format. Some tools directly integrate data from (1) Enterprise Resource Planning (ERP) systems, (2) databases, or (3) Comma Separated Value (CSV) files, all three in a proprietary way. However, developed in 2010, XES became the IEEE standard for "achieving interoperability in event logs and event streams" [11]. Today, XES is supported by the majority of process mining tool vendors.

XES defines three basic objects: log, trace and event. Log (the process) contains a collection of traces (execution instances) and a trace contains a collection of events [20]. Each object can contain an arbitrary set of strongly typed attributes in the form of key-value pairs. Every attribute value has a data type, like string, boolean, or date. To add semantics to these data types, XES defines the concept of extensions. An extension defines a set of attributes, their types, and keys with a specific semantic meaning.

3 Materials and Methods

This section describes which standards and tools were used in building the analytics suite and how we utilized and extended them to enable process mining based on HL7 FHIR.

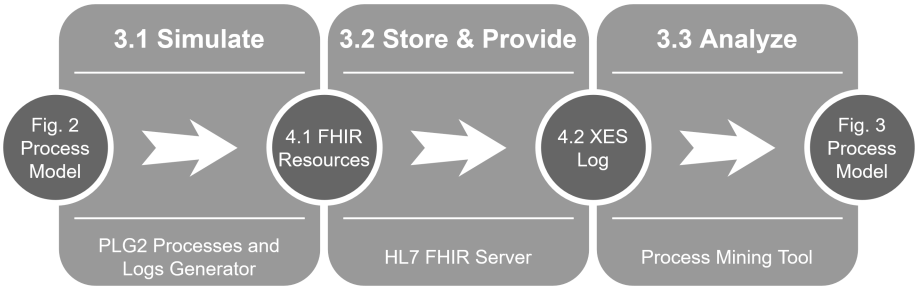


Fig. 1. The three steps of the interface test setting including the respective consumed and produced data. The numbers correspond to sections or figures in this paper.

Figure 1 depicts the three steps (1) simulate, (2) store&provide, and (3) analyze, that aim to show how the open standardized process analytics approach works. The circles represent data consumed and produced in those three steps.

To test the approach, a simple process was used. Figure 2 shows a simplified process model for an examination in a radiology practice using Business Process Model and Notation (BPMN). It shows the main steps from the appointment scheduling to the distribution of the diagnostic report. It is based on the work of Erickson et al. on business analytics in radiology [5] and on the process model used for evaluation in [7]. This is of course just an example and the approach is applicable to other healthcare domains as well.

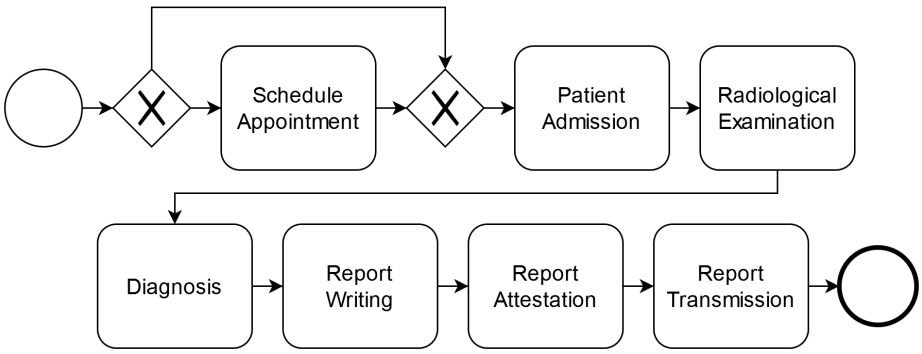


Fig. 2. BPMN process model of the radiology practice workflow based on [5, 7].

In the first step, a patient that, e.g., received a referral for a radiological examination, calls the practice to *schedule an appointment*. In some cases of our simulation, this step can be skipped and the patient arrives without a scheduled appointment. On the day of the examination, the patient arrives at the reception and is placed on the waiting list (*patient admission*). When called, the patient enters the procedure room and the *radiological examination* takes place. Afterwards, the radiologist makes a *diagnosis* and dictates the report. The *report writing* is done by trained specialists. The resulting report is *attested* by the radiologist. Finally, the report is sent to a requesting physician or handed out directly to the patient (*report transmission*).

3.1 Simulate

In order to be able to automatically generate process data, some sort of process engine or simulator is required. Burattin [2] developed a tool specifically designed to simulate processes and generate event logs for process mining, the Processes and Logs Generator (PLG2). The tool allows to generate and simulate random BPMN models, and to add randomized noise (e.g., double activity execution, skipping activities, etc.). The tool also allows to load an existing model, in our case the model from figure 2, and simulate it.

To use PLG2 for the simulation, we needed to make REST calls to our HL7 FHIR server. PLG2 allows to specify the execution time of different activities using Python scripts [2]. We adapted those scripts to execute REST calls using Client for URLs (cURL). By default PLG2 provides a single parameter, that is, the case identifier (caseId), to these python functions. We used this parameter to make the process instances distinguishable by deriving resource identifiers from it (i.e., patientId and encounterId).

Each activity in the process from figure 2 was extended with REST calls, creating, reading, or updating resources and executing operations on the FHIR server (according to the mapping described in the next section). The process was then simulated 10 times without randomized noise, each run resulting in one process instance recorded on the server.

3.2 Store & Provide

We set up a FHIR server including the required extensions and operations to automatically record audit trails, and to transform and provide this information in the XES format for process mining.

FHIR Server. We implemented our FHIR Server based on the open-source project "HAPI-FHIR Starter"[5]. This project provides a fully working FHIR server, including a database connection, based on the HAPI FHIR JPA project. Adjustable configuration files and the interceptor framework [19] create high

[5] https://github.com/hapifhir/hapi-fhir-jpaserver-starter

flexibility for custom changes and for adding extensions to the existing server implementation.

We utilized the Consent Interceptor, which amongst other functionalities has the ability to hook into the point of the server code, where a CRUD operation (e.g., creating an appointment or reading a patient record) has been finished. One of the Consent Interceptor's roles is to write audit trail records, creating an AuditEvent resource every time an operation has been finished successfully or with a failure.

In addition to the interceptor implementation, we provided the FHIR operation *$fhirToCDA* as part of our custom extensions to the server implementation. The operation can be executed on a specific instance of the DiagnosticReport resource and it returns an empty document to the client. An AuditEvent recording the execution of this operation in the context of a radiology workflow encounter will, for mapping purposes, be interpreted as a report transmission activity.

To query for an event log in the XES format, we extended our FHIR server by the *$xes* operation, which is defined to work on the AuditEvent resource type and is there to identify and transform all AuditEvents of the radiological workflow "rad-wf" into the XES format:

```
GET [fhirserver]/AuditEvent/$xes?plandefinition=PlanDefinition/rad-wf
```

Extending AuditEvent. We filled the AuditEvent resource with request details that are automatically provided for any standard CRUD operation. In order to be able to query for relevant AuditEvent resources, we needed to identify grouping elements. We decided to extend the AuditEvent resource by references to the Encounter and PlanDefinition resources (cf. section 5.1). Geared to the other resources containing the Encounter resource reference as part of their standard FHIR resource definition, we named the extended AuditEvent element "encounter". An additional extension "basedon" is used to reference the PlanDefinition resource "rad-wf", that defines the radiological workflow. This element can later be used to filter AuditEvent resources related to the executions of the radiological workflow process, while Encounter references are used to distinguish the single process instances (i.e., the traces).

Mapping FHIR AuditEvent to XES. For the test setting, we base our mapping on the assumption that Encounter identifier can be utilized as trace identifiers and that recorded events refer to a common process description, i.e., a medical guideline or pathway defined as a PlanDefinition. Of course, this is just one perspective, and different perspectives can be taken on the data (cf. section 5.3).

Let R be the set of all resources on the FHIR server. Let $A \subseteq R$ be the set of all AuditEvent resources, and $E \subseteq R$ be the set of all Encounter resources, and $P \subseteq R$ be the set of all PlanDefinition resources. All three subsets are disjoint, i.e., $A \cap P = \emptyset$, $A \cap E = \emptyset$, and $E \cap P = \emptyset$. Resources can refer to other resources

Table 1. Mapping table of operations on specific FHIR resources to activities of the radiology practice workflow, ordered by occurence in the simulated model in figure 2.

Operation	**FHIR Resource**	$\mapsto$	**Activity**
create	Appointment		Schedule Appointment
update	Appointment		Patient Admission
create	Procedure		Radiological Examination
create	Media		Diagnosis
create	DiagnosticReport		Report Writing
update	DiagnosticReport		Report Attestation
execute	*$fhirToCDA		Report Transmission

via the predicate refersTo$(r, r') :\Leftrightarrow (r, r') \in R$, where r' is referenced by r, i.e., r contains the identifier of r'.

Let $p_w \in P$ be the PlanDefinition resource "rad-wf" defining the radiology workflow. Then, $A_w = \{a \in R | \forall_{a \in A} \text{refersTo}(a, p_w)\}$ is the set of all AuditEvent resources recorded during the execution of radiology workflows.

For our mapping, let A_w be a set of disjoint sets A_{wi}, where every A_{wi} represents a set of AuditEvents recorded during a specific radiology workflow encounter $\exists e \in E$ of one patient. Then, every A_{wi} will be mapped to a trace σ in an XES event log L.

For testing the approach, we only map to mandatory fields in L, e.g., concept:name of the event (providing the activity name) and time:timestamp of the event (for ordering). Table 1 describes which recorded combination of operation and resource is mapped to which activity name. The timestamp is mapped directly from the recording time AuditEvent.recorded.

3.3 Analyze

Querying the FHIR server for AuditEvent resources using the $xes operation returns an XES event log. Since the operation already utilizes XES standard extensions (i.e., Concept and Time), the semantics of the fields are clear for process mining tools. The next step is to analyze if the simulated process matches the one stored and provided by the HL7 FHIR server. Thus, we want to compare the input model with a model generated based on the retrieved XES event log. We use the process mining tool ProM 6.9 [20] with the Visual Inductive Miner plugin [14] to generate a model.

4 Results

This section shows three exemplary results of the implementation: (1) a FHIR resource generated by the simulator, (2) the corresponding event in the XES event log, and (3) the process model created based on the event log. All results and examples can also be found in our GitHub open-source project[6].

[6] https://github.com/fhooeaist/ProcessMiningOnFHIR/

4.1 FHIR Resources

As described in the mapping in table 1, the Report Writing activity is associated with creating a DiagnosticReport resource. The simulator thus executes the following cURL statement:

```
POST [fhirserver]/DiagnosticReport
{ "resourceType": "DiagnosticReport",
  "subject": { "reference": "Patient/[patientId]" },
  "encounter": { "reference": "Encounter/[encounterId]" },
  "status": "preliminary",
  "code": {
    "coding": [ {
        "system": "http://loinc.org",
        "code": "LP31534-8",
        "display": "Study report"
    } ]
  }
}
```

This triggers the creation of an AuditEvent resource. This one is shown in abbreviated form, focusing on the elements relevant for the mapping:

```
{ "resourceType": "AuditEvent",
  "extension": [
    { "url": "https://fhirserver.com/extensions/auditevent-encounter",
    "valueReference": { "reference": "Encounter/[encounterId]" }},
    { "url": "https://fhirserver.com/extensions/auditevent-basedon",
    "valueReference": { "reference": "PlanDefinition/rad-wf" }}
  ],
  "action": "C",
  "recorded": "2020-08-14T08:42:51.523+02:00",
  "entity": [ {
    "what": { "type": "DiagnosticReport" },
    "detail": [ {
      "type": "RequestedURL",
      "valueString": "[fhirserver]/DiagnosticReport/"
    } ]
  } ]
}
```

The created AuditEvent resource refers to the respective Encounter resource and to the PlanDefinition resource "rad-wf" that defines the radiology workflow. The *action* field indicates the type of operation (C=Create) and the *entity* element contains details about the manipulated resource, i.e., the DiagnosticReport.

4.2 XES Log

The query for AuditEvent resources with the $xes operation returns the following XES event log (only one trace with one event is shown, extensions left out):

```
<?xml version="1.0" encoding="UTF-8" ?>
<log xmlns="http://www.xes-standard.org/">
  <string key="concept:name" value="PlanDefinition/rad-wf"/>
  <trace>
    <string key="concept:name" value="Encounter/enccase55"/>
    <event>
      <string key="concept:name" value="Report Writing"/>
      <date key="time:timestamp" value="2020-08-14T08:42:51.523+02:00"/>
    </event>
  </trace>
</log>
```

This detail of the resulting XES log shows the concept:name attributes on log and trace level, derived from the referenced PlanDefinition and Encounter resources respectively. The event (*report writing*) was generated for the AuditEvent resource presented in the previous section 4.1, that recorded the creation of a DiagnosticReport.

4.3 Process Model

Figure 3 shows the resulting model after importing the XES event log in ProM and analyzing it with the Inductive Visual Miner [14]. It is split up in two parts and highlights the similarity to the input model in figure 2. All traces were identified based on their Encounter reference and all AuditEvents were correctly mapped according to table 1. All 10 recorded executions are visible, with 5 skipping the first (*schedule appointment*) activity.

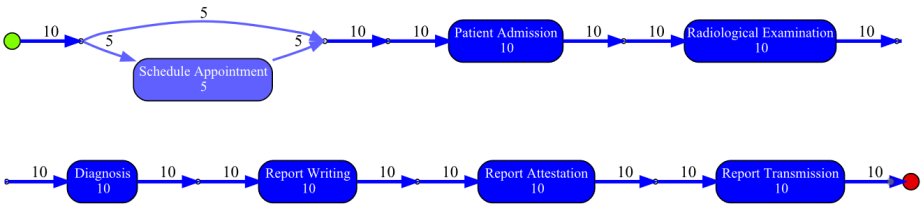


Fig. 3. Process model generated with the Inductive Visual Miner.

5 Discussion

The presented work is a proof of concept, making the case for a standards-based process analytics approach and making sure that the standard in development, HL7 FHIR, is aware of the capabilities and requirements of process mining. We were able to show how only minor extensions, namely the addition of Encounter and PlanDefinition references, and a simple mapping, enabled the analysis of the radiology practice workflow with process mining tools.

5.1 Impact on Standardization

In the FHIR Workflow project, the authors made a case for checking the usability of FHIR resources for process mining. Together, the working group members proposed the addition of a trace identifier to the AuditEvent and Provenance resources[7]: "We want to be able to search on all events (creates, updates, deletes, etc.) that happened during a given encounter, that happened based on a particular protocol or as a result of a particular order.". Based on the discussions in that group, we decided to use PlanDefinition and Encounter for the grouping and mapping approach. A proposal to extend AuditEvent to support this is currently under review for inclusion in the next FHIR release R5.

5.2 AuditEvent vs. Provenance

In this work we analyzed AuditEvent resources, building on existing approaches that aimed to analyze audit data [3, 7, 16]. However, HL7 FHIR also makes use of the concept of *provenance*, recording "information about entities, activities, and people involved in producing a piece of data or thing, which can be used to form assessments about its quality, reliability or trustworthiness" [6]. A Provenance resource is created by the client (i.e., the person or system conducting the work) as opposed to the AuditEvent resource, which is created automatically by a server. The client should explain for what purpose a resource was edited (created, updated, deleted). In addition, a client can add information about the process (or policy) behind the edit, and provide reasoning why something was done (i.e., which path of a process model was taken). However, Provenance is (1) not widely used (yet), and (2) not documenting non-changing access to a resource (i.e., read). To summarize, Provenance can provide more detailed information on a process, but relies on the clients to record it and might thus be not present at all. Further research on the utilization of the Provenance resource for process mining is needed.

5.3 Considering Different Perspectives

In our example, A_w, the set of all AuditEvent resources recorded during the execution of a radiology workflow (as defined by the referenced PlanDefinition "rad-wf"), was split to traces based on the referenced Encounter resources. However, in fact, A_w represents a multiset of traces, that can be split based on the perspective you take on the data. A more generic approach should thus indicate the grouping behaviour in the query, based on the concepts developed in [12].

Another viable perspective would be, for example, to look at the active participants of the workflow. AuditEvent.agent is described as "an actor taking an active role in the event or activity that is logged" [10]. Mapping name and role to the corresponding fields of the XES Organizational extension allows for additional analysis, e.g., social networks or handover of work for medical or care personnel.

[7] https://jira.hl7.org/projects/FHIR/issues/FHIR-28100

References

1. van der Aalst, W.: Academic view: Development of the process mining discipline. In: Process Mining in Action: Principles, Use Cases and Outlook, pp. 181–196. Springer (2020)
2. Burattin, A.: Plg2: Multiperspective process randomization with online and offline simulations. In: BPM (Demos). pp. 1–6 (2016)
3. Cruz-Correia, R., Boldt, I., Lapão, L., Santos-Pereira, C., Rodrigues, P.P., Ferreira, A.M., Freitas, A.: Analysis of the quality of hospital information systems audit trails. BMC medical informatics and decision making **13**(1), 84 (2013)
4. Erdogan, T.G., Tarhan, A.: Systematic mapping of process mining studies in healthcare. IEEE Access **6**, 24543–24567 (2018)
5. Erickson, B.J., Meenan, C., Langer, S.: Standards for business analytics and departmental workflow. Journal of digital imaging **26**(1), 53–57 (2013)
6. Groth, P., Moreau, L.: W3C Working Group Note "PROV-Overview" (2013), https://w3.org/TR/2013/NOTE-prov-overview-20130430/, last access: 1.8.2020
7. Helm, E., Paster, F.: First steps towards process mining in distributed health information systems. Int. Journal of Electronics and Telecommunications **61**(2), 137–142 (2015)
8. HL7 International: FHIR Specification (v4.0.1: R4) Executive Summary (2019), https://hl7.org/fhir/summary.html, last access: 1.8.2020
9. HL7 International: FHIR Specification (v4.0.1: R4) Operations (2019), https://hl7.org/fhir/operations.html, last access: 1.8.2020
10. HL7 International: FHIR Specification (v4.0.1: R4) Resource AuditEvent - Detailed Desc. (2019), https://hl7.org/fhir/auditevent-definitions.html, last access: 1.8.2020
11. IEEE Task Force on Process Mining: IEEE Standard for eXtensible Event Stream (XES) for Achieving Interoperability in Event Logs and Event Streams. IEEE Std. **1849-2016**, 53–57 (2016). https://doi.org/10.1109/IEEESTD.2016.7740858
12. IHE IT Infrastructure Technical Committee: Add restful atna (query and feed). IHE ITI-TF Supplement, Rev. 3 (2019)
13. IHE Radiology Technical Committee: Standardized operational log of events (sole). IHE RAD-TF Supplement, Rev. 1.2 - Trial Implementation (2018)
14. Leemans, S.J., Fahland, D., van der Aalst, W.: Process and deviation exploration with inductive visual miner. BPM (Demos) **1295**(8) (2014)
15. Mandel, J.C., Kreda, D.A., Mandl, K.D., Kohane, I.S., Ramoni, R.B.: SMART on FHIR: a standards-based, interoperable apps platform for electronic health records. Journal of the American Medical Informatics Association **23**(5), 899–908 (2016)
16. de Murillas, E.G.L., Helm, E., Reijers, H.A., Küng, J.: Audit trails in openslex: paving the road for process mining in healthcare. In: Int. Conference on Information Technology in Bio- and Medical Informatics. pp. 82–91. Springer (2017)
17. Rebuge, Á., Ferreira, D.R.: Business process analysis in healthcare environments: A methodology based on process mining. Information systems **37**(2), 99–116 (2012)
18. Rojas, E., Munoz-Gama, J., Sepúlveda, M., Capurro, D.: Process mining in healthcare: A literature review. Journal of biomedical informatics **61**, 224–236 (2016)
19. Smile CDR Inc.: HAPI FHIR Interceptors: Overview (2020), https://hapifhir.io/hapi-fhir/docs/interceptors/interceptors.html, last access: 25.8.2020
20. Verbeek, H., Buijs, J.C., Van Dongen, B.F., van der Aalst, W., et al.: Xes, xesame, and prom 6. In: CAiSE Forum. vol. 72, pp. 60–75. Springer (2010)

Deriving a sophisticated clinical pathway based on patient conditions from electronic health record data

Jungeun Lim[1], Kidong Kim[2], Minsu Cho[3], Hyunyoung Baek[2], Seok Kim[2], Hee Hwang[2], Sooyoung Yoo[2], and Minseok Song[1]

[1] Department of Industrial and Management Engineering, Pohang University of Science and Technology, Pohang, South Korea
[2] Office of eHealth Research and Businesses, Seoul National University Bundang Hospital, Seongnam, South Korea
[3] Research Institute of Industry & SME Strategy, Korea Institute of Industrial Technology, Seoul, South Korea

Abstract. Clinical pathway (CP), a standardized treatment process based on a clinical guideline, is widely used to reduce costs while maintaining or improving patient care quality. However, there is a gap between the actual clinical process and the guideline, that causes CP application to be disturbed. A study on developing a data-driven automated clinical pathway to obtain insight into real clinical processes has been conducted. Still, patient characteristics and conditions, which could cause a variation, have not been fully considered. In this study, we aimed to develop a framework to derive a sophisticated clinical pathway from electronic health records (EHRs) data by exploring process variations according to the patient characteristics and conditions. To validate the applicability of the proposed framework, We conducted a case study using the Total Laparoscopic Hysterectomy (TLH) CP data, which was retrieved from an EHR system of a tertiary general hospital in South Korea between January 2012 and April 2016. We found that diabetic TLH patients show different medical performances with other TLH patients. We developed a tailored CP that adds eleven orders over the standard TLH CP, and experts evaluated it as meaningful.

Keywords: Clinical pathways · electronic health records(EHR) · statistical analysis · evidence-based approach · clinical features · Business process analysis

1 Introduction

A clinical pathway (CP) is a standardized care process in a specific setting such as a particular surgery [7, 4]. The use of CPs is gaining interest to help decrease hospital costs and improve the quality of medical services by reducing undesired practice variability [13, 12]. Additionally, CPs shorten the length of hospital stays, lower costs, reduce complications and lower mortality [13, 8]. As

such, more than 80% of hospitals in the United States adopted CPs in the late 1990s [14], and currently, the implementation of CPs is widely contemplated by hospitals all over the world [21].

The traditional approach for developing a CP relied solely on the knowledge of clinical experts and clinical guidelines. Although the approach was a valuable method derived from solid theoretical backgrounds, it was limited by the time and effort required and the lack of generalization [19, 17]. Due to the highly dynamic, highly complex and ad hoc features of the medical treatment process, there is also a gap between the actual clinical process and the CP. As such, an automated approach from data is needed, and researchers have tried to resolve these challenges using process mining and data mining.

Mans et al. [10] applied heuristic miner, and a further work [11] used fuzzy miner and trace clustering to obtain insights from CPs. Huang et al. [4] proposed a new approach for mining CP patterns with time information from chronicle mining. Rebuge et al. [16] suggested a framework to compare the discovered CP and its variants using sequence clustering. Xu et al. [18] developed a more straightforward CP using the Latent Dirichlet Allocation technique. Additionally, researchers have employed further data mining techniques to develop CPs, such as frequent itemset mining [15], sequential pattern mining [15], and a rule induction algorithm [5].

These studies have contributed to developing the automated and accurate CPs based on data, deriving a standardized CP for the majority of patients. Despite these efforts using the data-driven approaches, it is still challenging to apply and complete CP with little effort in practice. In general, most hospitals only implement a single universal CP for a specific surgery or procedure. But, given the various clinical features of diabetes, cardiovascular, age, and medical history, a single CP cannot cover all different patients even with the same surgery; thus, a CP needs to be subdivided according to the clinical features. Therefore, with the aim of the increase of practical use, it is required to implement an approach for CP segregation with clinical features.

This study aims to identify the distinctive clinical characteristics that affect to distinguish a new clinical pathway. To this end, this paper suggests a framework consisting of four phases: data preparation, feature engineering, statistical analysis, and CP development. We first define the outcome measures and explanatory variables from the data. The matching rate, which represents a similarity between clinical trace and reference CP, is adopted as one of the medical performances for process-oriented assessment. Then, statistical testing is conducted to identify the key features highly related to clinical performance measures. Based on decisive factors from the statistical results, we distinguish a new CP (i.e., CP development) after post-hoc analysis with trace alignment. To validate the proposed framework, we performed a case study with real data from a tertiary hospital in South Korea.

The remainder of this paper is organized as follows. Section 2 explains the proposed framework. Section 3 shows a case study, and Section 4 discusses the results. Finally, Section 5 concludes the paper with future work.

2 Proposed Framework

In this section, we propose a framework for CP segmentation by patient characteristics. As shown in Fig. 1, the framework consists of four phases: data preparation, feature engineering, statistical analysis, and post hoc analysis & CP development. Data preparation, the first phase, aims to identify the data that can be utilized for data analysis by wrangling the collected data. Then, dependent (i.e., outcomes) and explanatory variables (i.e., patient characteristics) are defined in the feature engineering step. The statistical analysis phase conducts experiment to identify the relationship between outcome and independent variables. Lastly, in the post hoc analysis and CP development phase, we distinguish the new CP based on the result of comparing the clinical orders by statistical analysis and trace alignment.

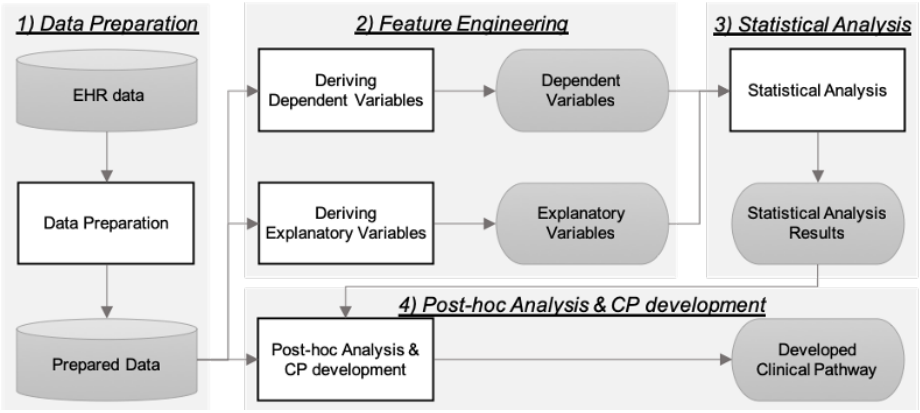


Fig. 1. The proposed framework in this paper.

2.1 Step 1: Data preparation

The first phase of the framework aims to prepare data with a suitable format for statistical analysis by collecting and pre-processing records. Clinical data generally are complex and heterogeneous [3]. There are four kinds of quality issues: missing data, incorrect data, imprecise data, and irrelevant data [1]. Missing data indicates that data is missing from logs, while incorrect data signifies that information recorded is not correct. Imprecise data represents that the level of data is too coarse, whereas irrelevant data means that information is not related at all with the log. These four types of quality issues are explicitly connected with the healthcare environment, and it needs to be processed thoroughly. To resolve these issues, users can choose proper data repair and noise removal methods based on the data quality. In our case, the most of issues was relevant with

missing data, and we tried to remove all problematic data. Details will be given in the Result section.

2.2 Step 2: Feature engineering

One of the main parts in our framework is to identify the patient characteristics that are highly relevant to the outcomes. To this end, we perform feature engineering to build a research model before the data analysis. As such, the second phase aims to derive dependent and explanatory variables implied for statistical modelling. In more detail, dependent variables represent the outcomes, such as the length of stays or matching rate, i.e., an indicator that signifies the difference between the clinical pathway and relevant clinical log [20], while independent variables signify the patient characteristics. They are derived by selecting or refining records from the prepared data.

Dependent variables (Outcome measures) Dependent variables represent the materials to evaluate the outcomes, such as length of stays, hospital costs, the amount of antibiotics used, and matching rates with respect to efficiency and complication rates, re-hospitalization rates, and mortality with respect to quality of the clinical services. Among these variables, in this study, we only employed the length of stays and matching rates, i.e., the efficiency-focused, because of the insufficiency of data related to the quality perspective. More in detail, we were not able to collect the patients' records who re-visited the hospital with the same diagnosis within the 30 days (i.e., re-hospitalization) or were turned out to be dead (i.e., mortality).

The length of stays is one of the critical indicators in most hospitals because it lowers the risk of infection and medical costs for patients. In this study, we derived the length of stays by calculating the difference between the admission date and discharge date.

The matching rates signify how patient records collected from the logs coincide with the orders in the CPs. Thus, the rates can be used to evaluate the practical application of the CP in the quantitative approach. The matching rate is formalized as follows [20].

$$\textit{CP order matching rate} = \frac{1}{2}(1 - \frac{M_{cp}}{N_{cp}}) + \frac{1}{2}(1 - \frac{R_{log}}{N_{log}}) \tag{1}$$

M_{CP} is the number of orders included in the CP but not shown in the log, N_{CP} is the number of orders included in the CP, R_{log} is the number of orders included in the log but not shown in the CP and N_{log} is the number of orders included in the log.

Explanatory variables (Patient characteristics) As introduced earlier, explanatory variables represent the materials that classify patients with their characteristics. Thus, regarding these characteristics, patients can be divided into

groups. For example, patients are divided into age groups, such as infants, children, young adults, middle-aged adults, and older adults. Additionally, they may be classified by whether they have a specific history or not.

EHR system contains numerous patient characteristics, including age, sex, family history, past history, and they can be categorized into three types: background information, clinical events, and non-clinical events. The background patient information signifies historical records of patients before hospitalization. This group includes age, sex, allergy, operation history, medication history, family disease history, and chronic diseases (diabetes, hypertension, hyperlipidaemia, and cardiovascular and cerebrovascular diseases). The second group is the data derived from the clinical events during hospitalization, such as transfer of wards, transfer of departments, diagnosis from another department (not from obstetrics and gynaecology), and operation from another department. The last category is related to the administrative information during hospitalization, including severity, admission type, Diagnosis Related Group (DRG).

2.3 Step 3: Statistical analysis

This step performs a statistical analysis to identify the distinctive patient characteristics considered for CP development. To this end, hypothesis testing is performed based on dependent and independent variables derived in Step 2. Regarding hypothesis testing, different types of methods are utilized considering the number of groups and shape of distributions. In this study, we applied two types of statistical analysis methods: Mann-Whitney U test and Jonckheere-Terpstra test.

Mann-Whitney U test The Mann-Whitney U test identifies whether two populations are equal or not [9]. As such, the test was applied when the patients were divided into two groups by a patient characteristic, such as sex and severity. Its null(H_0) and alternative(H_1) hypotheses are as follows.; H_0: Two populations are equal, H_1: Two populations are not equal.

Jonckheere-Terpstra test As a substitute for the Mann-Whitney U test, the Jonckheere-Terpstra test is applied when the number of groups is more than two (i.e., three or more) and they tend to increase or decrease [6]. For example, the changes of outcome variables can be identified by the increase in the number of operations. Letting di be the median for the population i, the null and alternative hypotheses are defined as follows; $H_0 : d_1 = d_2 = d_3 = \cdots = d_k, H_1 : d_1 \leq d_2 \leq d_3 \leq \cdots \leq d_k$ (where, k is the number of groups).

2.4 Step 4: Post hoc analysis & CP development

The last step compares the selected patients' clinical orders based on their characteristics and derives a new CP. Here, the critical patient characteristics are

employed from the statistically significant factors in Step 3. In this phase, patients are grouped by a specific feature, and the application rates of clinical orders are measured for each group. Then, the difference in the application rates of the orders between groups is identified. For example, if the order applies only to 90% of the severely ill group and 10% of non-severe patients, the order should be included in the CP of the severely ill group. Then, if a group of features differentiates multiple clinical orders, some traces from each group are sampled to visualize the differences and discuss with clinical experts. CP segmentation is performed when the clinical expert concludes that the functional group needs a new CP.

3 Case study

3.1 Introduction

A general tertiary hospital in South Korea has developed and applied numerous electronic CPs based on clinical experience to provide appropriate medical services to patients. In this case study, we primarily analyzed the Total Laparoscopic Hysterectomy (TLH) CP, which has been in use since August 2009. From the hospital's EHR system, log data of patients determined as candidates to be applied to the TLH CP were extracted from January 2012 to April 2016, resulting in data collected from 1100 inpatients. EHR data of patients' demographics, hospitalization, applied CP, surgery, diagnosis, transfers, referrals, physician orders including medications and labs, and CP history was extracted.

3.2 Data Preparation

Based on the collected data from 1100 inpatients, we performed data preprocessing. Among the four types of data quality issues, e.g., missing data, incorrect data, imprecise data, and irrelevant data, our data included the first type as we lacked the medical history of patients, such as operations and medication history. Additionally, the second-hand data collected from surveys, such as drinking and smoking, had many blank spaces. As such, those characteristics were removed from the data to be analysed. Furthermore, part of the clinical orders had incorrect data, such as an unexpected hold (3.4%) and immediate removal by systems (2.5%). These were also excluded, and finally, the data was prepared.

3.3 Feature engineering

Dependent variables (Outcomes) As introduced earlier, we applied the length of stays and matching rates as dependent variables (i.e., outcomes). Regarding the length of stays, the average value was 4.57 days (median: 4 days and standard deviation (SD): 1.8 days). Regarding the matching rate, the average was 0.716 (median: 0.724 and SD: 0.053).

Explanatory variables (Patient characteristics) After preparing the data, we selected 11 explanatory variables based on a thorough discussion with clinical experts: diabetes, hypertension, hyperlipidaemia, cardiovascular, cerebrovascular, severity, operations, transfers of departments, transfers of wards, diagnosis from other departments (not from obstetrics and gynaecology), and referrals to other departments.

Only a small number of patients had chronic diseases, including diabetes, hypertension, hyperlipidaemia, cardiovascular, and cerebrovascular at 3.5%, 4.7%, 1.5%, 0.1%, and 0.6%, respectively. The number of patients with severity, however, was relatively high at 33.1%. Regarding the number of operations, most patients received only one operation while 0.9% of patients received two operations. Regarding transfers of departments, only four patients (0.4%) changed departments. Lastly, regarding the other characteristics (e.g., transfers of wards, diagnosis from other departments, and referrals to other departments), for each feature, more than 50% of the patients were not associated with the feature at all, but the remaining patients had more than one frequency.

3.4 Statistical analysis

Among the 11 independent variables (i.e., patient characteristics), only six, e.g., diabetes, hypertension, severity, transfers of wards, diagnosis from other departments (not from obstetrics and gynaecology), and referrals to other departments, were considered for statistical testing because the sample size for testing should be sufficient (i.e., more than 30) [9], and the sample sizes for the other features are not sufficient.

We applied two different statistical testing methods: the Mann-Whitney U test and Jonckheere test. The Mann-Whitney U test was applied to diabetes, hypertension, and severity while the Jonckheere test was employed for the remaining variables. Table 1 presents the statistical testing results of the length of stays and matching rates on patient characteristics.

Table 1. Statistical testing results on patient characteristics.

Patient characteristics	p-value		Test type
	LOS	Matching rates	
Diabetes	< 0.01	< 0.01	Mann-Whitney U test
Hypertension	0.014	0.045	Mann-Whitney U test
Severity	< 0.01	< 0.01	Mann-Whitney U test
Transfers of wards	< 0.01	0.035	Jonckheere test
Diagnosis from the other departments (not from obstetrics and gynaecology)	< 0.01	0.149	Jonckheere test
Referrals to other departments	< 0.01	< 0.01	Jonckheere test

As a result of the statistical tests, diabetes, severity, transfers of wards, diagnosis from other departments and referrals to other departments significantly

affected the length of stays while the matching rates were significantly affected by diabetes, severity, and referrals to other departments. Therefore, we concluded that only three features, e.g., severity, diabetes, and referrals to other departments, are key characteristics for CP segmentation.

Based on these results, we had a thorough discussion with clinical experts. First, regarding severity, we determined that the result was caused by incorrect application of the CP in cancer patients, not the CP target patients. In the hospital, clinicians sometimes applied the CP to cancer patients because there was no significant difference in clinical operation processes between the two. The cancer patients, however, required a longer stay and different routines from the CP patients. Thus, we determined that it was misleading that there was an impact on clinical outcomes. Additionally, regarding the referrals to other departments, the domain experts concluded that the feature needs to be managed by monitoring rather than CP development. For these reasons, we performed further post hoc analysis and CP development based on diabetes.

3.5 Post hoc analysis & CP development

Considering diabetes, we analyzed the differences in clinical orders between diabetic and non-diabetic patients. The total number of diabetic and non-diabetic patients was 38 and 1062, respectively. We performed trace alignment to visualize how the order records of each group differ. For simplicity, in each group, 20 patients, who stay in the hospital for four days, are sampled, and the result of trace alignment is in Fig.2.

Additionally, we employed the CP development methodology [2], which derives an optimal set of clinical orders that maximize the matching rates. Based on the exploited method, we received clinical orders for diabetic and non-diabetic patients. After, the developed CP for diabetes was compared with that for non-diabetes. We identified that 11 clinical orders, e.g., Pot chloride, Humalog, Palonosetron, Ephedrine, Electrolyte panel, Glucose, DM diet (for diabetes), BST, Infusion pump, Interceed, and Simple hysterectomy, were applied for most of the diabetic application rates. In contrast, two clinical orders (i.e., Granisetron and Other dermatological) were utilized only for non-diabetic patients.

Table 2 provides the clinical order application rates of diabetic and non-diabetic patients. Overall, we were able to identify a clear difference in each code's application rates by the group. Therefore, we concluded that the new CP for diabetes should be distinguished from the general one.

4 Discussion

The results of the analysis showed that diabetes affects medical outcomes, such as the length of stays and matching rates. To this end, we identified that glucose control is the reason for the extended hospital stays and the lower matching rates. Patients with diabetes require a specific amount of time to control their

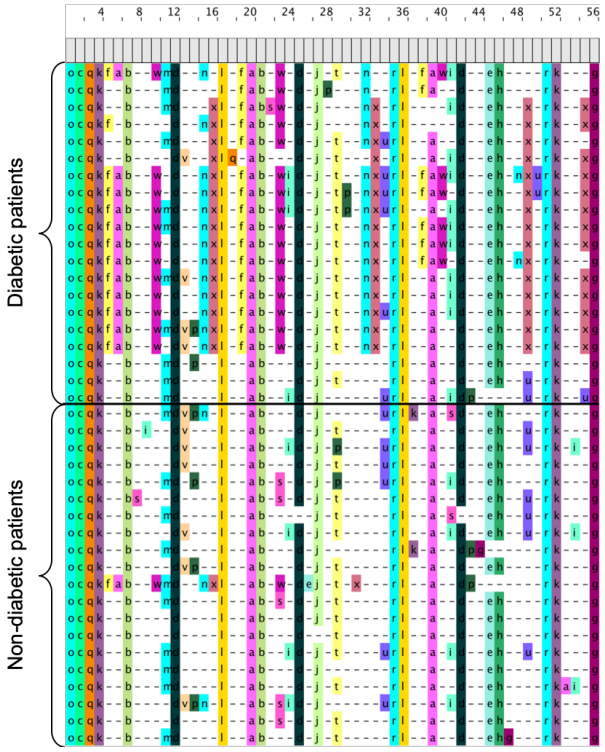


Fig. 2. Trace alignment result of diabetic patients and non-diabetic patients.

blood sugar before surgery, which can lead to longer hospital stays. Additionally, the diabetic patients received surgery later than general patients.

Regarding the lower matching rate for diabetic patients, we found that controlling the patient's blood sugar affected the results through the post hoc analysis. We identified that diabetic patients received insulin (e.g., humalog) with Alberti regimen and dextrose fluid (e.g., pot chloride) containing potassium chloride to ensure adequate water, electrolyte, and feeding before operations. Additionally, diabetic patients received tests to check blood sugar and electrolytes for glucose control. Moreover, some materials (e.g., infusion pump) were also utilized for diabetic patients to inject the proper medicines. Therefore, we determined that these orders are required entirely for diabetic patients with both data and clinical perspectives.

This research has important contributions for both practice and research standpoints. As far as practical use is concerned, this research helps to develop the clinical decision support system by resolving the large demands from hospitals to continuously improve and manage CPs. Despite the facts that hospitals generally cannot develop and enhance CPs due to an insufficient workforce, time, and costs, however, it is required to implement a tool that gives accurate clinical

Table 2. Clinical order application rates of diabetic and non-diabetic patients.

Order Information		Application rates (%)	
Type	Name	Diabetic	Non-diabetic
Medications	Pot Chloride	84.2	2.7
	Humalog	84.2	2.0
	Palonosetron	57.9	47.1
	Ephedrine	50	36.3
	Granisetron	42.1	56.2
	Other Dermatologicals	44.7	49.3
Lab Test	Electrolyte panel	79.0	9.5
	Glucose	79.0	3.2
Diet	DM diet (for diabetes)	71.1	1.6
Treatment	BST	89.5	2.2
Procedures	Infusion Pump	81.6	7.1
	Interceed	57.9	47.6
	Simple Hysterectomy	50	39.9

pathways to clinicians, driving to provide high-qualified patient-centric services. In this standpoint, this paper is of value as it automatically recommends distinctive patient characteristics and develops a new CP with a data-driven approach.

Also, as far as the research standpoint is concerned, this paper is different from existing works that merely discover a one-off CP and provides a direction that enables the continuous development of improved CPs with a statistical approach. Furthermore, the patient characteristics and clinical outcome measures derived in this research are applicable to multiple clinical research disciplines, such as real-time monitoring and prescriptive analytics in hospitals.

Despite these contributions, this paper has some challenges. First, there has been a problem that the number of patients to be analyzed is reduced because latest data of short-term period data must be used to reflect the latest order information. Nonetheless, it is significant that we were able to segment the CP according to the patient condition of diabetes. The framework presented in this study considerably contributes in terms of managing the clinical pathway and practical use of the clinical pathway and will continue to demonstrate its usefulness through further data acquisition.

Also, this research did not address the inter-relationship between patient characteristics and thus only aimed at developing new CPs for each patient feature. However, it is possible to construct CPs that consider multiple patient characteristics at once (e.g., diabetic-female-TLH CP). Furthermore, we limited clinical outcome measures to length of stays and matching rates. Future studies should be expanded to more scalable methodologies, including patient costs and the use of antibiotics. Lastly, the analysis result presented in this paper was only based on a single hospital. As there are differences in CPs and data between hospitals, the study may lack generalizability. Thus, we need to perform more case studies using data from multiple hospitals. We believe that we can build a more robust framework for CP segmentation by resolving these issues.

5 Conclusion

In this paper, we proposed a framework for CP segmentation based on patient characteristics. In this process, we performed feature engineering to define the clinical outcome measures related to CPs (i.e., dependent variables) and patient characteristics (i.e., independent variables). We also conducted statistical testing using the Mann-Whitney U test and Jonckheere test, and finally a new CP was distinguished from the general CP.

This paper proposes guidelines to increase the applicability of CPs and suggests how to develop CP variants using patient characteristics and clinical outcomes. Additionally, the proposed framework has a distinctiveness that enables the continuous development of improved CPs different from existing works that merely discover a single CP. Therefore, we believe that our methodology is helpful for practical use.

In future studies, we will consider the inter-relationship between patient characteristics for CP segmentation. Additionally, other clinical outcomes, such as patient costs and the use of antibiotics, may be included. Furthermore, more case studies should be performed to validate our approach and make various use cases.

Acknowledgement

This work was approved (IRB No. B-1609/361-105) by the Institutional Review Board of the SNUBH, which waived patients' informed consent. This work was funded by the MSIT (Ministry of Science and ICT), Korea, under the ITRC (Information Technology Research Center) support program (IITP-2020-2018-0-01441) supervised by the IITP (Institute for Information & Communications Technology Promotion).

References

1. Bose, R.J.C., Mans, R.S., van der Aalst, W.M.: Wanna improve process mining results? In: 2013 IEEE symposium on computational intelligence and data mining (CIDM). pp. 127–134. IEEE (2013)
2. Cho, M., Kim, K., Lim, J., Baek, H., Kim, S., Hwang, H., Song, M., Yoo, S.: Developing data-driven clinical pathways using electronic health records: The cases of total laparoscopic hysterectomy and rotator cuff tears. International journal of medical informatics **133**, 104015 (2020)
3. Cios, K.J., Moore, G.W.: Uniqueness of medical data mining. Artificial intelligence in medicine **26**(1-2), 1–24 (2002)
4. Huang, Z., Lu, X., Duan, H.: On mining clinical pathway patterns from medical behaviors. Artificial intelligence in medicine **56**(1), 35–50 (2012)
5. Iwata, H., Hirano, S., Tsumoto, S.: Construction of clinical pathway based on similarity-based mining in hospital information system. Procedia Computer Science **31**, 1107–1115 (2014)

6. Jonckheere, A.R.: A distribution-free k-sample test against ordered alternatives. Biometrika **41**(1/2), 133–145 (1954)
7. Lenz, R., Blaser, R., Beyer, M., Heger, O., Biber, C., Bäumlein, M., Schnabel, M.: It support for clinical pathways—lessons learned. international journal of medical informatics **76**, S397–S402 (2007)
8. Macario, A., Horne, M., Goodman, S., Vitez, T., Dexter, F., Heinen, R., Brown, B.: The effect of a perioperative clinical pathway for knee replacement surgery on hospital costs. Anesthesia & Analgesia **86**(5), 978–984 (1998)
9. Mann, H.B., Whitney, D.R.: On a test of whether one of two random variables is stochastically larger than the other. The annals of mathematical statistics pp. 50–60 (1947)
10. Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., Van Der Aalst, W.: Process mining techniques: an application to stroke care. In: MIE. vol. 136, pp. 573–578 (2008)
11. Mans, R.S., Schonenberg, M., Song, M., van der Aalst, W.M., Bakker, P.J.: Application of process mining in healthcare–a case study in a dutch hospital. In: International joint conference on biomedical engineering systems and technologies. pp. 425–438. Springer (2008)
12. Newman, B.: Enhancing patient care: case management and critical pathways. The Australian journal of advanced nursing: a quarterly publication of the Royal Australian Nursing Federation **13**(1), 16 (1995)
13. Panella, M., Marchisio, S., Di Stanislao, F.: Reducing clinical variations with clinical pathways: do pathways work? International Journal for Quality in Health Care **15**(6), 509–521 (2003)
14. Pearson, S.D., Goulart-Fisher, D., Lee, T.H.: Critical pathways as a strategy for improving care: problems and potential. Annals of internal medicine **123**(12), 941–948 (1995)
15. Perer, A., Wang, F., Hu, J.: Mining and exploring care pathways from electronic medical records with visual analytics. Journal of biomedical informatics **56**, 369–378 (2015)
16. Rebuge, Á., Ferreira, D.R.: Business process analysis in healthcare environments: A methodology based on process mining. Information systems **37**(2), 99–116 (2012)
17. Weiland, D.E.: Why use clinical pathways rather than practice guidelines? The American journal of surgery **174**(6), 592–595 (1997)
18. Xu, X., Jin, T., Wei, Z., Wang, J.: Incorporating topic assignment constraint and topic correlation limitation into clinical goal discovering for clinical pathway mining. Journal of Healthcare Engineering **2017** (2017)
19. Yang, W., Su, Q.: Process mining for clinical pathway: Literature review and future directions. In: 2014 11th International Conference on Service Systems and Service Management (ICSSSM). pp. 1–5. IEEE (2014)
20. Yoo, S., Cho, M., Kim, S., Kim, E., Park, S.M., Kim, K., Hwang, H., Song, M.: Conformance analysis of clinical pathway using electronic health record data. Healthcare informatics research **21**(3), 161–166 (2015)
21. Zhang, Y., Padman, R., Patel, N.: Paving the cowpath: Learning and visualizing clinical pathways from electronic health record data. Journal of biomedical informatics **58**, 186–197 (2015)

Exploration with Process Mining on How Temperature Change Affects Emergency Services

Juan José Lull[1][0000−0002−4399−950X], Onur Dogan[2][0000−0003−3543−4012], Angeles Celda[4], Jesus Mandingorra[4], Lenin Lemus[1], Miguel Ángel Mateo Pla[1], Javier F. Urchueguía[1], Gema Ibanez-Sanchez[1][0000−0003−1824−281X], Vicente Traver[1][0000−0003−1806−8575], and Carlos Fernandez-Llatas[1,3][0000−0002−2819−5597]

[1] Instituto Universitario de Tecnologías de la Información y de las Comunicaciones (ITACA), Universitat Politècnica de València, Camino de Vera S/N,Valencia 46022, Spain
{jualulno, geibsan, vtraver, cfllatas}@itaca.upv.es
[2] Izmir Bakircay University, Department of Industrial Engineering, Izmir, 35665, Turkey
onur.dogan@bakircay.edu.tr
[3] Department of Clinical Science, Intervention and Technology (CLINTEC) Karolinska Institutet, Stockholm, Sweden
carlos.fernandezllatas@ki.se
[4] Consorci Hospital General Universitari de València, Av. de les Tres Creus, 2, 46014 València, Spain

Abstract. The way patients are treated in Emergency Services changes during the year, depending on many factors. One key component is weather temperature. Some seasonal maladies are tightly related to temperature, such as flu in cold weather or sunburn in hot weather. In this study, data from a hospital in Valencia was used to explore how harsh weathers affect the emergency service, obtaining information about probable impacts of global warming in healthcare systems. Some illnesses, such as heat stroke, are more prevalent with higher temperature, but more interestingly, they also take more time to attend the patients. Rapid changes in temperature are also analyzed through Process Mining techniques.

Keywords: Process mining · Emergency · Weather conditions · Healthcare system.

1 Introduction

Emergency departments (EDs) work seven days, 24 hours a week. They are key departments that provide urgent care to the patients. Since many patients get further care after the ED's first response, they are regarded as the gateway to other hospital departments. The EDs aim to present urgent care to treat people

recover from their illnesses or at least alleviate the symptoms. Well-performed and standard processes can accomplish this aim in the ED, where healthcare professionals collaborate systematically. The increasing number of patients causes a crisis of agglomeration in the gateway of hospitals [14]. Although it is well-known among professionals and literature that most EDs are frequently crowded, many questions wait for their answers [3]. Among these questions, one might be how global warming affects emergency services.

Intergovernmental Panel on Climate Change (IPCC) points out that weather conditions will probably become more hotter or colder frequently and intensely in the following years [9, 12]. Large parts of the World, especially Asia, Europe and Australia have encountered the increased recurrence of heatwaves [10]. Besides, human mortality rates related to extreme hot weather have raised with global warming [10]. Several reasons may affect the correlation between disease and global warming, such as local demographics, economic welfare, underlying disease risk, weather variability in seasons, and available [7]. Another reason is that steep changes in daily temperature may have an impact on ED processes. For this consideration, more reliable intellection of disease conditions during temperature changes is an essential tool for health practitioners and the investigation of ED processes is gaining more and more attention [5, 8].

Despite progress in the analysis of ED processes, novel strategies are required for complexity, diversity and non-adaptability reasons [2, 11]. ED processes are not adjustable and adaptable from another process model because of their nature and complexity. This complexity makes it hard to provide a clear representation of the patient flow. Hence, most investigations focus on the observations to discover the process model, which is time-consuming and unreliable. Process models are the central part of crowded ED problems. Therefore, they should represent real and reusable patient flows to find acceptable solutions. Process mining (PM) automatically creates process models using real data stored in the IT system as event logs [13]. By applying PM methods, the actual ED processes followed by patients can be discovered to see the effects of global warming.

The studies that are presented in the next section, show the relation between higher temperatures and extra attention time, and explore the connection between patient cases and harsh, sudden temperature changes. This shows the potentiality of using PM in the study of global warming and healthcare.

2 Case Study

Data were available from 483,229 visits to the Emergency Service at a Hospital in the city of Valencia. These were records from the years 2015 till 2018. The records included:

- Patient ID
- Date and time of arrival to the service
- Date and time of the start of the triage and its end.
- Waiting queue assigned to the patient in the triage.

- The specific service that attended the patient (e.g. surgery, dermatology) and timestamp, both at the beginning and end of the attention to the patient.
- Patient destination (e.g. home, hospitalization, another medical service).
- Patient's gender
- Patient's date of birth
- Patient diagnostic

Daily temperature information was also available, including per day:

- Average temperature,
- Minimum temperature,
- Maximum temperature.

Across the years, subjects usually go to the hospital more than once. Specifically, 192,884 patients generated the 483,229 visits, with an average of 2.5 visits to the hospital per patient.

2.1 Assigning Temperature to Cases and Discretization

With the help of the PALIA suite [4], daily temperature information about Valencia city was fused with the Emergency Service data, assigning temperature to the date of each case.

Temperature information was then discretized, generating sub-groups of cases: 15-20°C, 20-25°C, 25-30°C, 30-35°C.

Inaccurate data (i.e. blank information, wrong dates) were removed, leaving 393,963 correct traces corresponding to visits to the ED.

PALIA process mining algorithm was applied to the data.

The Interactive Process Indicator (IPI) with information from all the visits may be seen in Figure 1, where the green-to-red color gradient represents shorter to longer duration in the nodes (accounted for by median duration) and lower to higher number of visits in the transitions.

The Waiting 1 to 5 and Attention 1 to 5 nodes represent the queues (and level of emergency, top to bottom) that each patient is assigned after the arrival, at the triage step. Afterwards, the patient returns home, though he or she could also be admitted into the hospital, or finish in *exitus*, among other possibilities. The emergency service modelling has been described elsewhere [6].

2.2 Temperature and Heat Strokes

The first study is related to daily weather temperature and heat strokes. As the WHO points [1], heat is one important factor that affects mainly the elder population, causing cardiovascular and respiratory diseases. In 2003, an excess 70,000 elderly people died due to a heat wave [1].

Interactive Process Indicators were generated for each temperature range and it was observed that the attention of patients with heat stroke took longer the higher the temperature.

As seen in Figure 2, attention time is highest for the group including daily temperatures of 25-30°C, while there were not enough cases in the 30-35°C range.

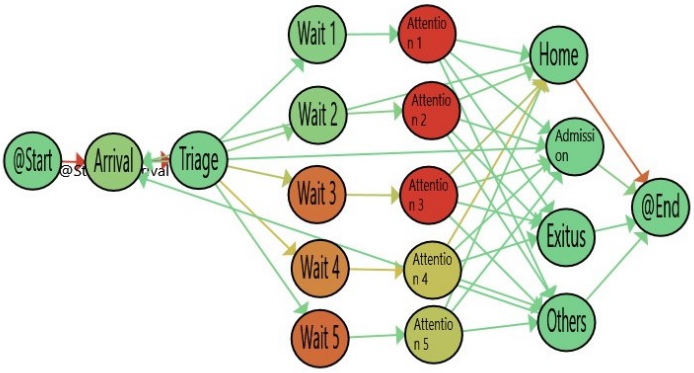


Fig. 1: IPI representing all the visits, including 393,963 traces after incorrect traces were discarded. Redder color in nodes represent higher time in that phase while redder color in transition means larger number of cases in that transition.

2.3 Otitis cases related to temperature

A high number of otitis (inflammation of the ear due to infection) cases were detected that related to high temperature. Their IPI is shown in Figure 3.

As can be seen in Figure 3, there were no otitis cases among the 568 diagnosed ones that were considered as serious, since none were triaged in the most urgent queues, 1 and 2.

It was observed that the number of otitis patients related to the number of general patients increased with temperature, as seen in Figure 4.

Although the attention of those patients took little time, their wait time was very high (as seen in Figure 3. This indicates a higher load of waiting rooms.

2.4 Harsh changes in temperature and ED

In order to study how sudden changes in weather (a phenomenon related to climate change) affect the ED, harsh changes in day-to-day temperature were detected and those with a higher sudden change were selected and their processes were visualized and studied.

In order to select the days with higher changes, they were compared: Each day's minimum temperature was subtracted to the minimum temperature from the previous day. The same calculation was performed for the maximum and the average temperatures. Finally, those values were multiplied and a threshold of 100 was introduced, accounting for 14 days with increased temperature and 17 days with reduced temperature. Two groups were created, one with harsh increases in temperature from day to day, and the other one with harsh decreases.

The IPI that represents days without steep day-to-day changes in temperature is shown in Figure 5.

The comparison between days with a high change in temperature compared to the previous one, generally showed higher attention times. Specifically, for increases in temperature, attention in queues 2 and 3 took longer than days with no significant change in temperature (see Figure 6). The same situation happened for steep temperature decreases, with a higher attention time and wait time for patients classified in queue number 3, and wait time for queue 5 was also higher, along with the triage time (see Figure 7). These findings were statistically significant. There were no statistically significant reductions in time.

In order to assess the significance in the differences, the normality of each population of durations was assessed by the Kolmogorov-Smirnov test. In case both populations are normal, a Student T test is applied. Otherwise, a Mann-Whitney-Wilcoxon is applied. In any case, a p value is applied as a threshold to determine statistical significance between the populations. In this study a p value of 0.05 was set as the threshold for statistical significance.

Yellow circles around nodes indicate a significant statistically difference between the duration in the population of days with a steep change compared to days without important changes in temperature. In the IPIs that compare to the baseline, greener means higher times while redder means lower (negative) times (Figures 6.b and 7.b).

3 Conclusion and Discussion

This study considered data collected from an emergency department (ED) at a Hospital in Valencia city, from 2015 to 2018. 483,229 visits created by 192,884 patients were investigated for four years. The effects of temperature on heat stroke and otitis cases are investigated in this study by analyzing process flows, along with a general investigation about the effects of steep changes in temperature on the Emergency Departments.

Valencia is a warm Mediterranean coastal city, so the weather is generally mild. However, we could detect changes in the processes inside the Emergency Department that depended on weather temperature.

Firstly temperature data was categorized and linked to process cases to explore possible effects of global warming. Then PALIA algorithm created process flows of patients under categorized (discretized) temperature data. It was observed that Heat Stroke processes in EDs took longer the higher the temperature. There were also many more cases in the range of 25 to 30°C. The number of cases confirms the intuition that sunburns are more prevalent the higher the temperature (it should be considered that there were roughly 26% days with an average temperature at or above 25°C across the years of the study). The higher treatment time span per case could be thought of as intuitive too, but in this case the effect of global warming on the EDs is clear: Since global temperature keeps increasing, sunburns are expected to grow in number and EDs will have more cases that will need extra attention time.

In the case of otitis, this was an unexpected finding. It should be reviewed how much confounding factors played a role in the cases, such as infections due to longer times spent in swimming pools as is usually the case during the summer vacations. This could nevertheless that waiting rooms could be overcrowded in humid areas with high temperatures.

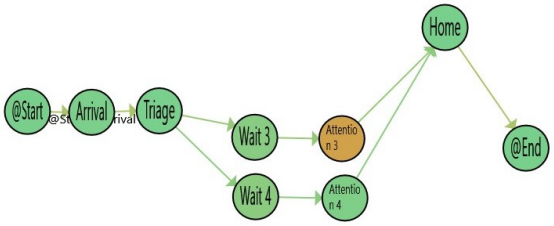
The study also presented the effects of sudden changes in weather conditions to the ED. Generally, time spent at the waiting rooms and while being attended were longer for both sudden temperature increases and decreases. This exploration points in the direction that the more the sudden changes in temperature, the more collapsed EDs will be. And sudden changes in temperature are more and more frequent due to climatic change.

With the presented results, the study puts forwards that global warming has a significant impact on Emergency Department processes.

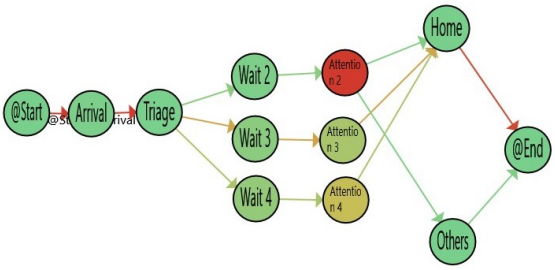
References

1. WHO Temperature Report. https://www.who.int/news-room/fact-sheets/detail/climate-change-and-health (2020), accessed: 2020-08-30
2. Abo-Hamad, W.: Patient pathways discovery and analysis using process mining techniques: an emergency department case study. In: International Conference on Health Care Systems Engineering. pp. 209–219. Springer (2017)
3. Bergs, J., Vandijck, D., Hoogmartens, O., Heerinckx, P., Van Sassenbroeck, D., Depaire, B., Marneffe, W., Verelst, S.: Emergency department crowding: Time to shift the paradigm from predicting and controlling to analysing and managing. International Emergency Nursing **24**, 74–77 (2016)
4. Fernández-Llatas, C., Benedi, J.M., García-Gómez, J., Traver, V.: Process mining for individualized behavior modeling using wireless tracking in nursing homes. Sensors **13**(11), 15434–15451 (2013)
5. Frumkin, H., Hess, J., Luber, G., Malilay, J., McGeehin, M.: Climate change: the public health response. American journal of public health **98**(3), 435–445 (2008)
6. Ibanez-Sanchez, G., Fernandez-Llatas, C., Martinez-Millana, A., Celda, A., Mandingorra, J., Aparici-Tortajada, L., Valero-Ramon, Z., Munoz-Gama, J., Sepúlveda, M., Rojas, E., Gálvez, V., Capurro, D., Traver, V.: Toward Value-Based Healthcare through Interactive Process Mining in Emergency Rooms: The Stroke Case. International Journal of Environmental Research and Public Health **16**(10), 1783 (may 2019)
7. Kinney, P.L., O'Neill, M.S., Bell, M.L., Schwartz, J.: Approaches for estimating effects of climate change on heat-related deaths: challenges and opportunities. Environmental science & policy **11**(1), 87–96 (2008)
8. Knowlton, K., Rotkin-Ellman, M., King, G., Margolis, H.G., Smith, D., Solomon, G., Trent, R., English, P.: The 2006 california heat wave: impacts on hospitalizations and emergency department visits. Environmental health perspectives **117**(1), 61–67 (2009)
9. Meehl, G.A., Tebaldi, C.: More intense, more frequent, and longer lasting heat waves in the 21st century. Science **305**(5686), 994–997 (2004)
10. Pachauri, R.K., Allen, M.R., Barros, V.R., Broome, J., Cramer, W., Christ, R., Church, J.A., Clarke, L., Dahe, Q., Dasgupta, P., et al.: Climate change 2014: synthesis report. Contribution of Working Groups I, II and III to the fifth assessment report of the Intergovernmental Panel on Climate Change. Ipcc (2014)

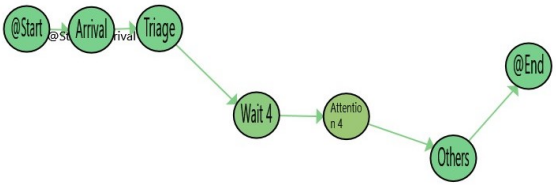
11. Rebuge, Á., Ferreira, D.R.: Business process analysis in healthcare environments: A methodology based on process mining. Information systems **37**(2), 99–116 (2012)
12. Stocker, T.F., Qin, D., Plattner, G.K., Tignor, M., Allen, S.K., Boschung, J., Nauels, A., Xia, Y., Bex, V., Midgley, P.M., et al.: The physical science basis. contribution of working group i to the fifth assessment report of the intergovernmental panel on climate change. Computational Geometry **18**, 95–123 (2013)
13. Van Der Aalst, W.: Process mining: data science in action. In: Process mining, pp. 3–23. Springer (2016)
14. Verelst, S., Pierloot, S., Desruelles, D., Gillet, J.B., Bergs, J.: Short-term unscheduled return visits of adult patients to the emergency department. The Journal of emergency medicine **47**(2), 131–139 (2014)



(a) 20-25°C



(b) 25-30°C



(c) 30-35°

Fig. 2: Interactive Process Indicators (IPIs) with groups of average temperature per day, in Heat Stroke patients.

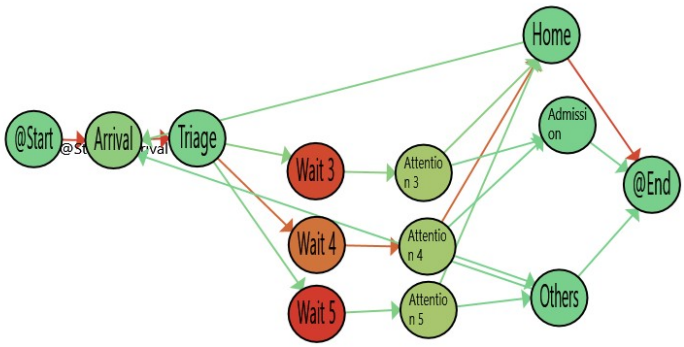


Fig. 3: IPI for otitis patients. Gradient colors represent the same durations as in Figure 1.

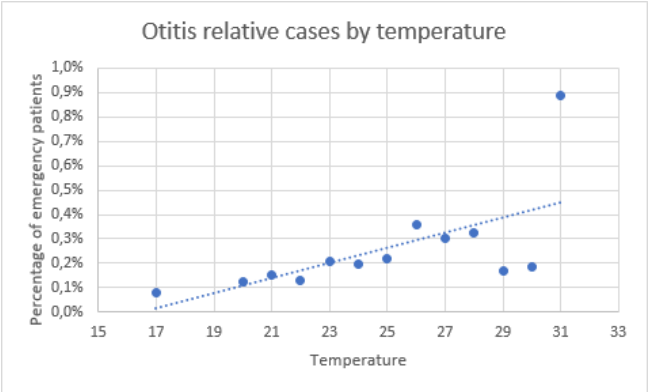


Fig. 4: Relative number of otitis cases (percentage of cases by general emergency patients), and its increase with temperature.

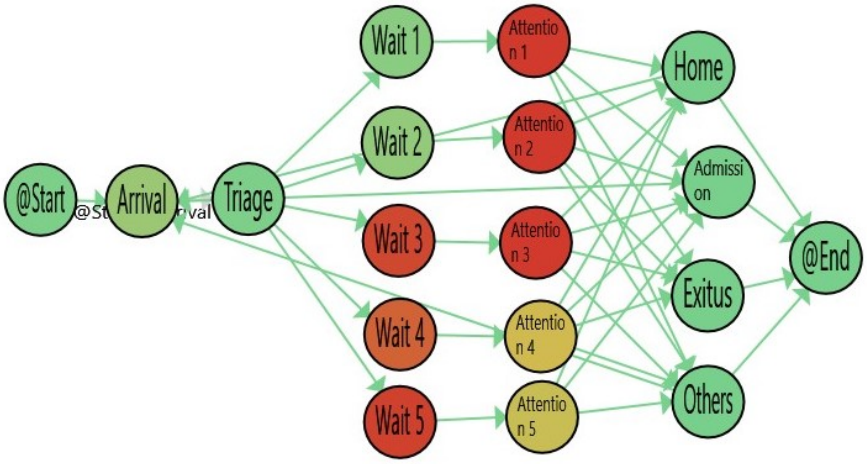
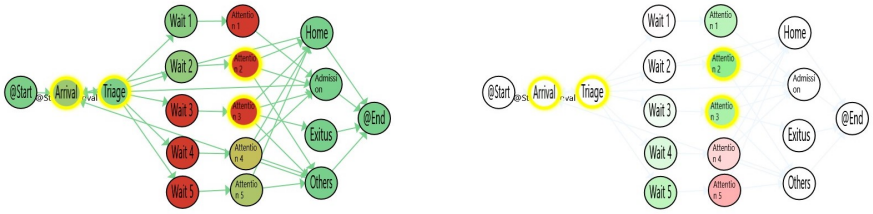
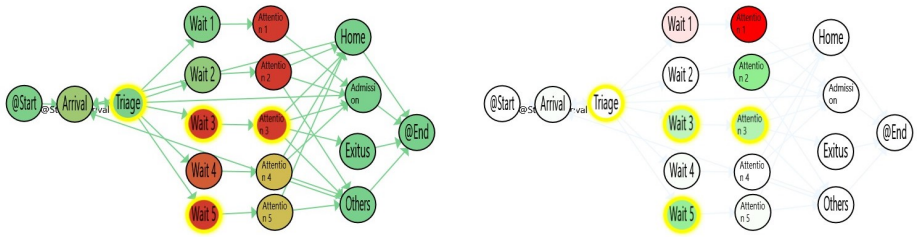


Fig. 5: IPI with information on days without harsh temperature changes. The redder the color, the longer the time at a node. Colors (thus durations, are directly comparable between this figure and the ones with harsh increases and decreases in temperature).



(a) IPI for harsh increase in temperature (b) Same IPI, compared to baseline

Fig. 6: IPIs for ED patients in days with steep increases in temperature compared to the previous day.



(a) IPI for harsh decrease in temperature (b) Same IPI, compared to baseline

Fig. 7: IPIs for ED patients in days with steep decreases in temperature compared to the previous day.