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.

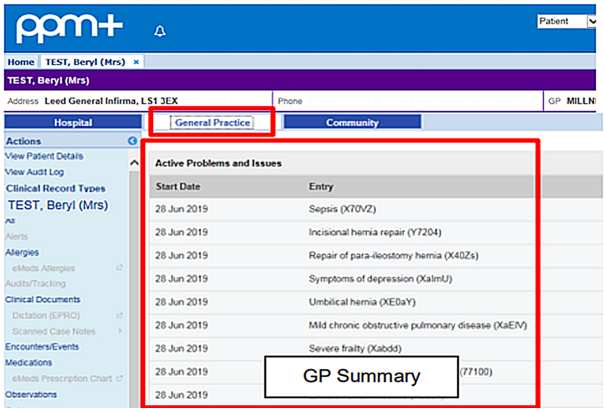

Fig. 1. Screenshot of the GPTab in the PPM EHR system, from the PPM support website [18].

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).

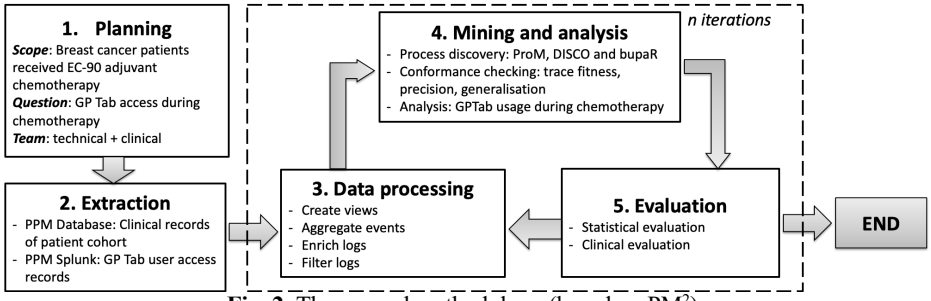

Fig. 2. The general methodology (based on PM^2)

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.

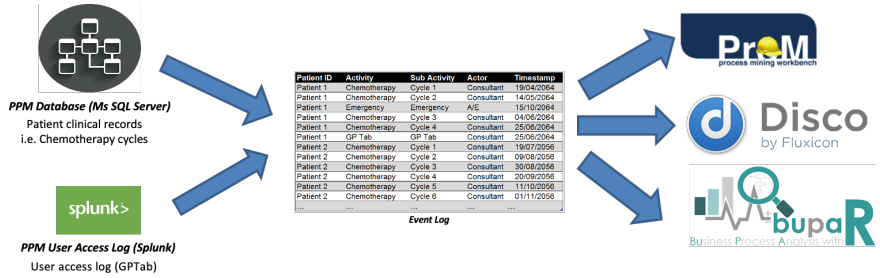

Fig. 3. The extraction stage, combining patient clinical records with user access records.

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.

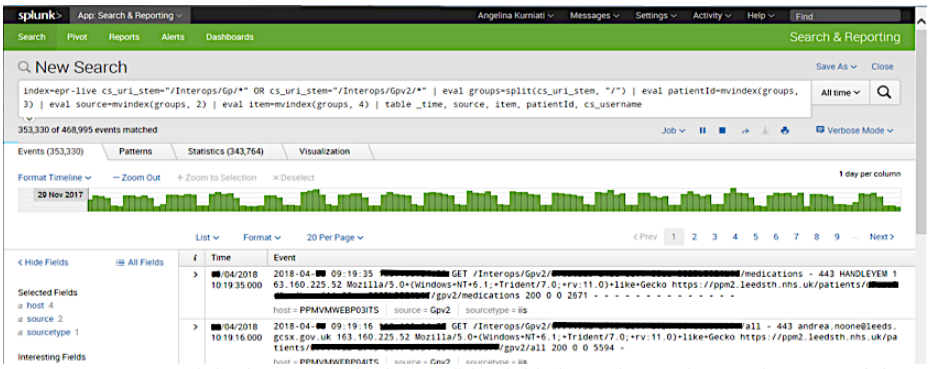

Fig. 4. A query result in the PPM Splunk. Confidential information such as Patient ID and dates are blocked in black.

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.

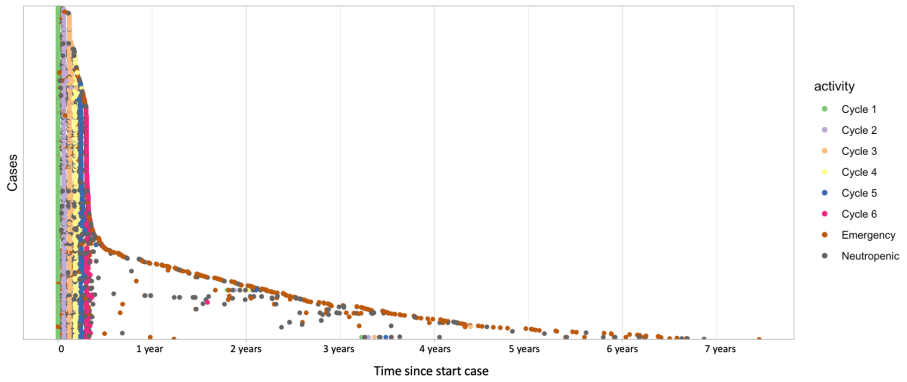

Fig. 5. Dotted chart showing adverse events during six chemotherapy cycles. The x-axis shows duration from the first activity to the last one. The y-axis shows patient id, sorted by durations.

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%).

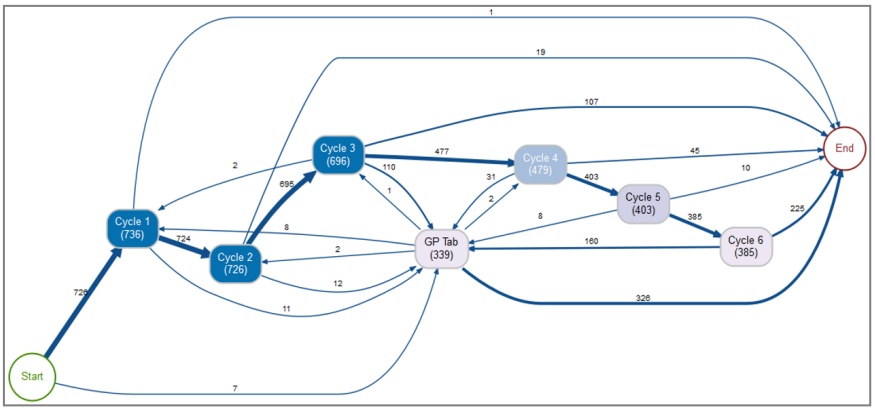

Fig. 6. Process model showing GPTab access during chemotherapy cycles (bupaR). It shows that GPTab was mostly accessed after *Cycle 3*, *Cycle 6*, or *Cycle 4.*

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.

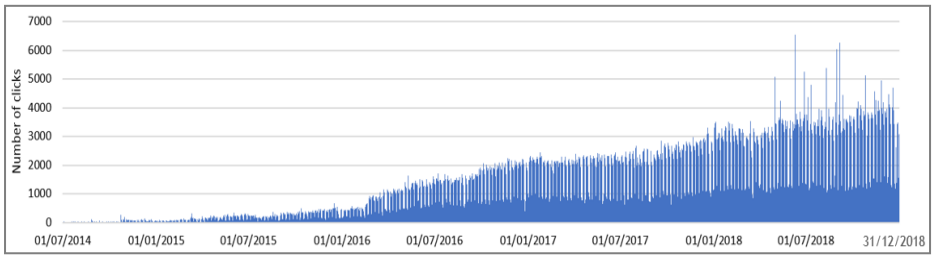

Fig. 7. GPTab clicks each day. It shows that the number of clicks generally increased over time, with steady fluctuations showing the pattern of weekday- and weekend- usages.

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.

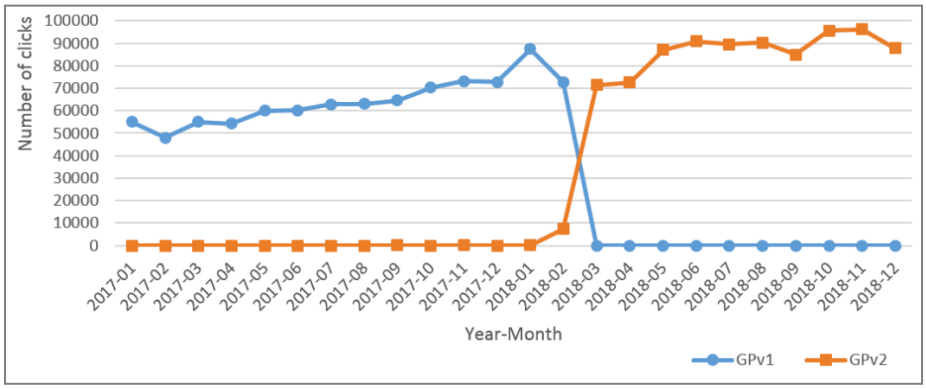

Fig. 8. Monthly usage of GPTab during 2017-2018. The blue dots are monthly usage of the first version (GPv1) and the orange dots are those of the second version (GPv2).

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.