Beyond X-Raying a Care-Flow: Adopting Different Focuses on Care-Flow Mining

Filip Caron, Jan Vanthienen, Jochen De Weerdt, and Bart Baesens

K.U.Leuven, Faculty of Business and Economics, Leuven Institute for Research on Information Systems (LIRIS), Naamsestraat 69, Leuven, Belgium
{Filip.Caron,Jan.Vanthienen,Jochen.DeWeerdt,Bart.Baesens}@econ.kuleuven.be

Abstract. Health-care processes are typically human-centric processes characterized by heterogeneity and a multi-disciplinary nature. The individual treatment of a patient is often unique, which makes process discovery a real challenge. Additionally, the informational value of a spaghetti-like process model is rather limited.

In this contribution we propose new applications of care-flow/process mining techniques that result in useful information. Two focuses will be further developed; the focus on department-based sub processes and the focus on a specific treatment/drug.

Key words: Process Mining, Care-flow, Health-Care, Business Process Intelligence Challenge

1 Introduction

While business process support for structured business processes (mainly for manufacturing and logistics) has always been an important research topic, the growing importance of service organizations (e.g. in health-care) with their human-centric processes has triggered the need for different approaches towards business process support [1]. Although the basic process might be straightforward, these processes, in general contain more flexibility, routings, loops, human judgment, freedom and variability than traditional processes.

Because of this need for process flexibility and variability, supporting systems for human-centric services typically allow much more degrees of freedom than standard administrative or production systems. Although this flexibility has important advantages such as the ability to make quick decisions based on recent developments, the increasing self-responsibility amongst employees, etc., this often results in poor insight and control over what really happens.

Typical examples of human-centric business processes include claim handling at an insurance company, social security decisions etc., but the focus for this

contribution will be placed on care-flows. For this purpose we use the event log of a care-flow process of the Department of Gynaecology at a Dutch Hospital as provided by the organizing committee of the First International Business Process Intelligence Challenge (BPIC'11) and that can be found on [2]. A preliminary analysis of the event log confirms the unstructured process assumption. Spaghetti-like process model are obtained with the heuristics miner plug-in [3] (see figure 1) and the performance sequence diagram analysis tool [4] indicates that there are virtually no repeating patterns. While trying to retrieve a comprehensible model for the patients evolution through this care-flow would be the most obvious challenge, this contribution aims at stimulating the development and adoption of different analysis approaches for unstructured processes and care-flows in particular. In the next paragraphs we will present two possible focuses:

– Analyzing **department-based sub processes**. We opted for the radiotherapy department as an example.
– Analyzing the use of a **specific therapy**. In this contribution we will further investigate the use of Paclitaxel, a treatment for cancer and other malignancies.

The analyses will be performed with ProM and its plug-ins [4].

Note that the conclusions drawn in this contribution are purely based on the behavior present in the event log and basic medical knowledge that can be found on the Internet. In order to use these conclusions the medical correctness and relevance must be determined by a professional.

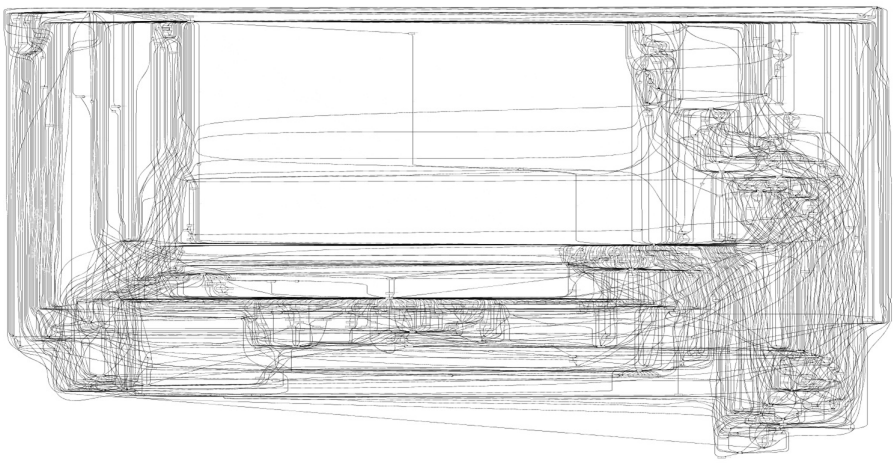


Fig. 1. Heuristic miner: resulting process model

2 Analyzing Department-Based Sub Processes

In general it can be concluded that most of the contemporary business processes are not isolated, instead they are interlinked and synchronized with other processes. By combining these related processes we obtain a super process that in addition to the entities' specific behavior, specifies the interaction between them. For example, the provided event log describes the care-flow as it is perceived in the Department of Gynaecology, however, the complete care-flow might have started with the family doctor and ended at the reconstructive plastic surgery department (e.g. breast reconstruction after a mastectomy).

Instead of moving up to the super process, we will further focus on a specific department present in the event log, namely Radiotherapy. This department treats all sorts of cancer and other malignancies (e.g. cervical malignancy or endometrial cancer) with basically three treatments (i.e. teletherapy, hyperthermia therapy and Brachytherapy).

2.1 Observation 1: Discovering the Process Model with Heuristics Miner

Placing the focus on Radiotherapy made it possible to obtain an understandable process model for the treatment of multiple and very different cancers or other malignancies from the perspective of this single entity (see fig 2).

In order to obtain this process graph we executed some (minor) data preprocessing steps:

- An event filter was placed to capture only the events originated at the Department of Radiotherapy.
- During the event selection phase we also decided to remove all events of the activity 'administratief tarief - eerste pol', as we assume that the sole purpose of these events is of administrative nature (e.g. invoicing) and therefore does not affect the health-care procedures of the department.
- The addition of an artificial start and end event for each process instance.
- All concepts of type 'behandeltijd - eenheid x - megavolt' were renamed 'behandeltijd', since we assume that all of these activities deal with determining the specific parameters of a radiation session/treatment.
- The radiotherapy options were rationalized, the process model distinguishes the following radiotherapy options:
  - Brachytherapy that covers the 'brachytherapie - interstitieel -bijzond' and 'brachytherapie - interstitieel - intensi'.
  - Hyperthermia therapy that contains the activities 'hyperthermie' and 'hyperthermie behandeling - h - 1'.

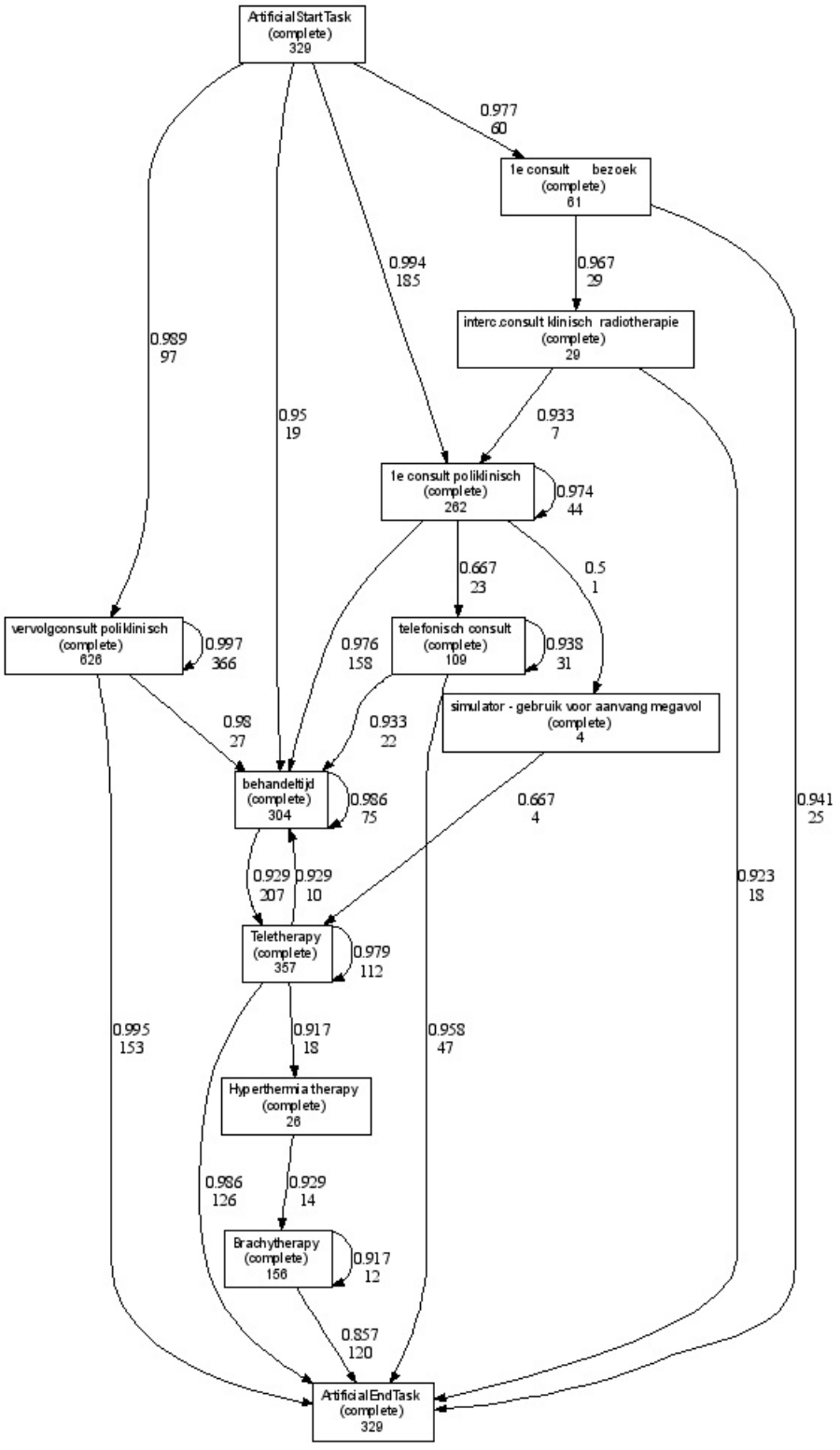


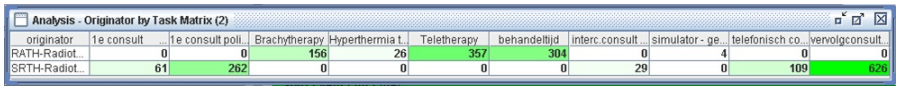
Fig. 2. Heuristic miner (standard parameterization): sub process model for the Department of Radiotherapy

– Teletherapy is the renamed ‘teletherapie - megavolt fontonen bestrali’ event.

The interpretation of the obtained process model is rather straightforward. Firstly, this care-flow sub process tends to start with some kind of a medical consult (i.e. seeing a physician) or if needed with obtaining multiple opinions of physicians at different levels of specialization. Besides initial consults (i.e. activities ‘1e consult bezoek’ and ‘1e consult poliklinisch’), we distinguish a transition between the ‘artificial start task’ and ‘vervolgconsult poliklinisch’ that might indicate that the full care-flow for these 97 patients was not retained in the event log. Followed by the determination of the parameters for the radiotherapy session/treatment in ‘behandeltijd’. Finally, the execution (sequence) of the different therapy options can be observed.

2.2 Observation 2: Layered structure of the Dutch health-care system

The filtered event-log clearly represents the layered structure of the Dutch health-care system, a distinction is made between second and third order medical care. Patients are stimulated to first contact a specialist at the polyclinic (i.e. producer code RATH in event log), where a physical examination and primary diagnose will be made. While in most cases the patients are referred to third order medical care (e.g. radiotherapy) provided by the academic hospital, there are process instances that end after this first contact. For example, in case 65 the process instance trace only includes a ‘1e consult bezoek’ or case 16 that solely consists of an execution of the activity ‘vervolgconsult poliklinisch’. This separation of concerns can be clearly observed in the originator-by-task matrix [4], see figure 3.


Analysis - Originator by Task Matrix (2)

originator	1e consult ...	1e consult poli...	Brachytherapy	Hyperthermia t...	Teletherapy	behandeltijd	interc.consult ...	simulator - ge...	telefonisch co...	vervolgconsult...
RATH-Radiot...	0	0	156	26	357	304	0	4	0	0
SRTH-Radiot...	61	262	0	0	0	0	29	0	109	626

Fig. 3. Originator-by-task matrix: separation of concerns

Excluding the process instances that start with a first contact at the polyclinic, uncovers 21 cases (or 6,5% of all cases in the filtered log) that do not follow this medical care hierarchy. Possible reasons might include: a first contact at a non-co-located polyclinic or a start event that occurred before the time-frame of this event log. However, an in-depth analysis of cases 23 and 107 might be required since they also lack the activity ‘behandeltijd’, which seems a more serious process deviation.

2.3 Observation 3: Links with Other Sub Processes in the Care-Flow

As previously mentioned sub processes are very interlinked, consequently it might be interesting to look at the interactions of the Department of Radiotherapy with other departments. The social network miner plug-in [5] in combination with the handover metric confirms this assumption, collaboration (and thus process synchronization) occurs with anesthesiology clinic, day center - ward, function center ENT, general lab clinical chemistry, internal specialism clinic, maternity ward, medical microbiology, nuclear medicine, nursing ward, obstetrics and gynaecology clinic, pathology, radiology and recovery room/high care. This can visualized with the analyze social network plug-in, see figure 4.

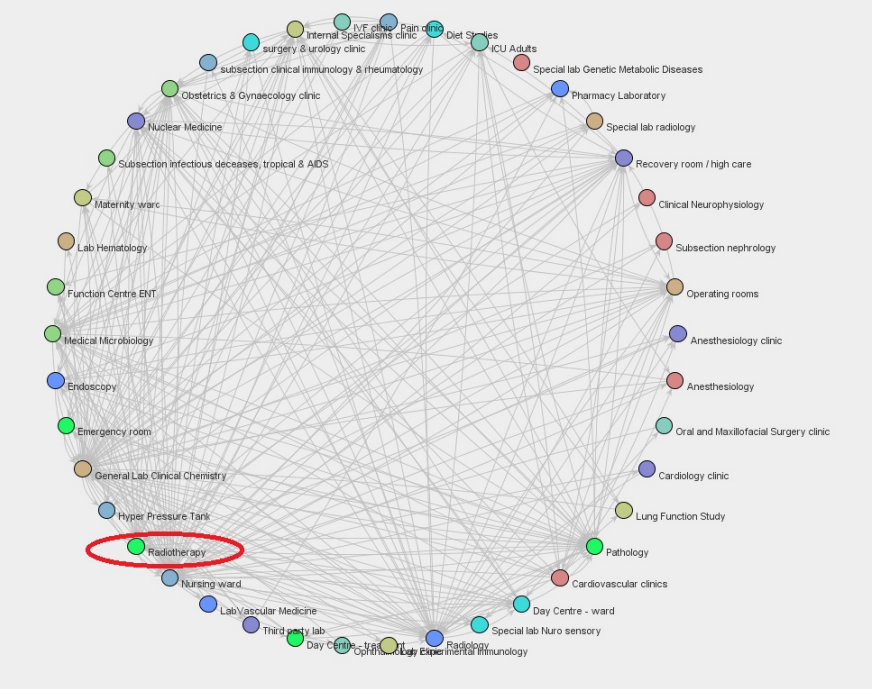


Fig. 4. Social network analyzer to reveal interactions with other departments

With the performance sequence diagram analysis plug-in (selecting options component type on the department field) [4] it can be concluded that there is no standard pattern of interaction, each pattern in which radiotherapy occurs has a frequency equal to 1.

2.4 Observation 4: Common Sequence between Treatments

The heuristics miner suggests a clear sequence between the treatments used at the Department of Radiotherapy, namely first teletherapy followed by hyperthermia therapy and finally Brachytherapy. This section further investigates this phenomenon with the LTL-checker plug-in [6]. In order to efficiently perform this analysis we first define an additional event filter, which filters out the events that do not correspond to a radiation therapy option.

Firstly, in 190 out of 215 cases teletherapy is the first treatment option that is selected. Amongst the remaining 25 cases we find for example case 27, in which a 'telefonisch consult' is performed followed by a 'vervolgconsult'. This might indicate that the patient has previously received some sort of treatment, but that it did not occur in the selected time-frame or at another hospital. We also remark that several cases do not include additional treatments and therefore we consider these people cured (e.g. case 124).

Secondly, is hyperthermia therapy the second treatment in the sequence of treatments? Testing the 'eventually_activity_A' reveals that only in a minority of the cases hyperthermia therapy is performed (i.e. 22 cases or approximately 10% of all cases that include at least one radiotherapy treatment). The 'eventually_activity_A_then_B' statement, with activity A representing teletherapy and activity B denoting hyperthermia therapy, returns true in 16 out of 22 cases (approximately 73%). An analysis of the incorrect process instances reveals that after a hyperthermia treatment a new teletherapy treatment has been performed, which might indicate the start of a new treatment cycle.

Thirdly, is the Brachytherapy usually the last treatment in a treatment cycle? This is true unless the person has been cured before. In 144 process instances the Brachytherapy occurs in the audit trail. For 119 of these cases (approximately 83%) it can be asserted that the Brachytherapy occurs after the execution of the teletherapy. In addition, from the 22 process instances that include hyperthermia therapy, only 15 of them also contain Brachytherapy. The statement 'eventually activity hyperthermia therapy then Brachytherapy' evaluates true for 14 out of 15 cases (approximately 93%).

In conclusion, most of the cases start with teletherapy and for those cases that contain brachytherapy there has been previously performed at least one teletherapy session. When a hyperthermia session is performed, it tends to occur between teletherapy and brachytherapy.

3 Analyzing the Use of Specific Therapies

For the second set of analyses we opted to focus on the use of Paclitaxel in the care-flow. The business object that is used as a trace in this process analysis is

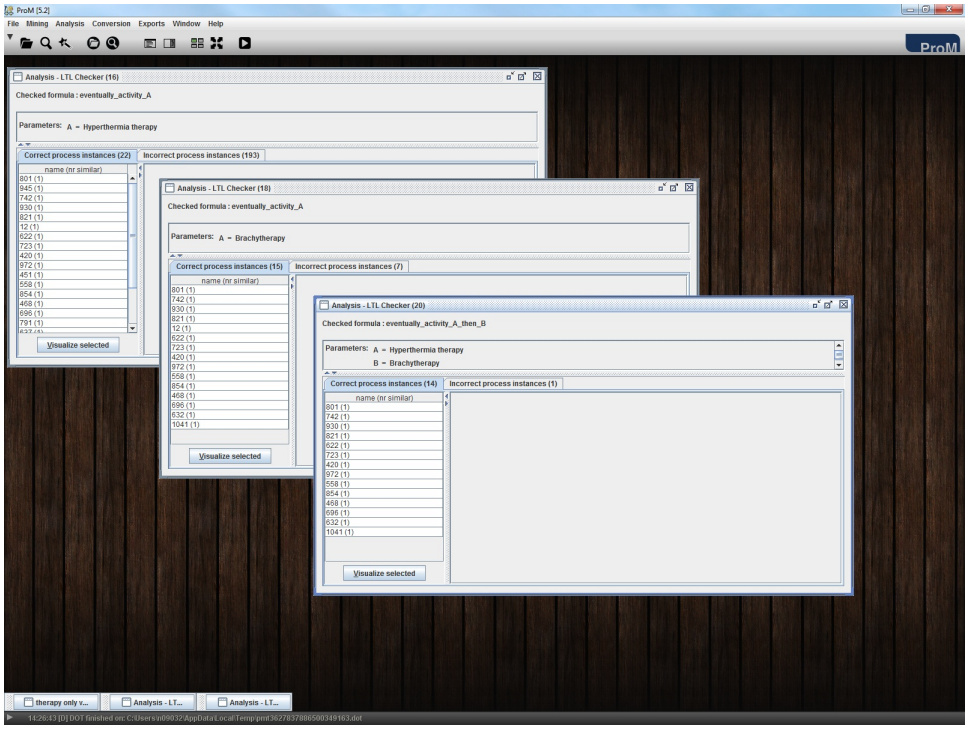


Fig. 5. LTL-checker for uncovering a sequence between treatments

also a patient. Paclitaxel is based on a mitotic inhibator and is used as a drug in cancer chemotherapy and Kaposi's sarcoma (i.e. dermatological condition). This treatment caught our attention because of its high values for 'number of executions'. In the next sections we present some observations on the behavior found in the event log.

3.1 Observation 1: AC-Paclitaxel Chemotherapy

Although several sources indicate the benefits of the sequential use of AC (i.e. doxorubicin and cyclophosphamide) and Paclitaxel, both in terms of chances for disease-free survival and for overall survival (e.g. [7, 8]), we were unable to find any patient who receives both the AC and Paclitaxel therapy. However, we did find that a minority of the cases in which Paclitaxel was used also report the use of Doxorubicin (through event: doxorubicine liposomal –caelyx–).

From the diagnosis descriptions provided in the event log we found it reasonable to assume that Paclitaxel was being used for cancer treatment, which is exactly the scope of the NICE Guidance TA108 [7].

Table 1. LTL-checker based results for AC-Paclitaxel Chemotherapy

	Paclitaxel used	**Paclitaxel only**	**Paclitaxel and Doxorubicin**
Absolute number	68	63	5
Relative numbers	100%	92,6%	7,4%
Case examples	12, 30, 796, 800, etc.	12, 30, 47, 77, etc.	119, 190, 209, 796, 810

3.2 Observation 2: Deviation between Prescribed Average Number of Cycles and the Real Average

In the NICE Guidance TA108 [7] the cost per patient is estimated at approximately £4000 for an average of four cycles of treatment. Based on the assumption that one case describes the full treatment of an individual person we will investigate if this number of average cycles is also applicable in Dutch hospital.

We were able to determine an average of approximately 7 treatment cycles and conclude that for approximately 30% of the patients 4 cycles or less was adequate, see figures 6 and 8. Consequently, the estimated cost per patient seems strongly underestimated. Further research must be performed in order to determine what caused the high deviation of the real number of treatment cycles compared to the prescribed number of treatment cycles, perhaps the lack of a combination with AC. Additionally, it would be interesting to analyze the periods between treatment cycles, as it appears that the optimal timing is every three weeks. However, without exact timestamps this analysis could not be performed.

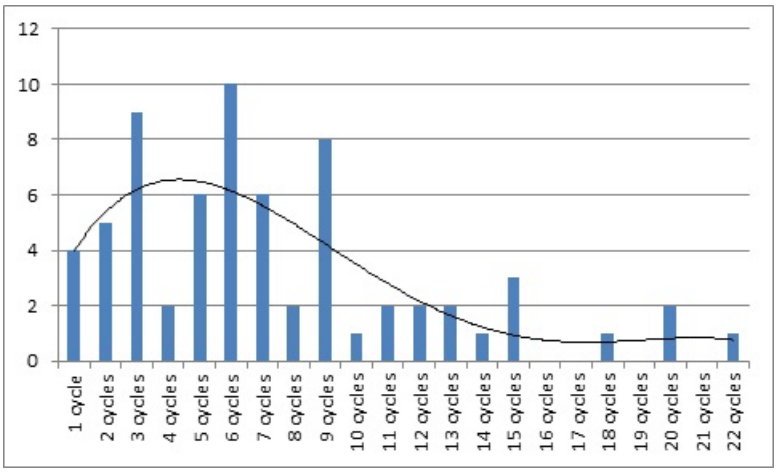


Fig. 6. Distribution of the number of treatment cycles per patient

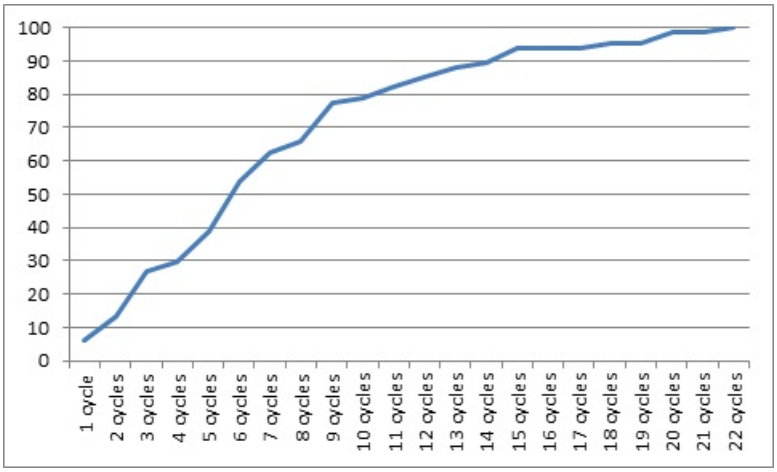


Fig. 7. Cummulative distribution of the number of treatment cycles per patient

3.3 Observation 3: Nurses Mostly Administer Paclitaxel

After filtering the event log until it only contains the Paclitaxel, an originator-by-task matrix [4] was created, see figure. Note the originator is determined by the two following fields 'producer code' and 'org:group'. This analysis clearly shows that the drug is most often administered by the a nurse and that normally no intervention of a medical doctor is needed.

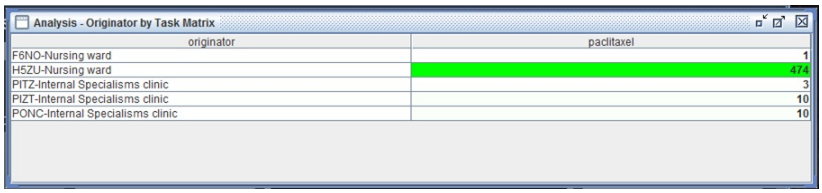


Fig. 8. Originator-by-task matrix for administering Paclitaxel

3.4 Observation 4: Versatile Use of Paclitaxel

According to the National Cancer Institute, Paclitaxel is approved by the Food and Drug Administration (FDA) to treat ovarian, breast cancer and AIDS-related Kaposi sarcoma [9]. A close inspection of the event log uncovers the use of Paclitaxel for the treatment of different kinds of malignancies at the ovary (e.g. treatment '3201' and diagnosis 'maligniteit ovarium'). However, the event log shows that the drug is also used for malignancies at the uterus (e.g. treatment '103' and diagnosis 'adenoca: corpus uteri st Illa'), the endometrium (e.g.

treatment '3301' and diagnosis 'maligniteit endometrium') and the cervix (e.g. treatment '113' and diagnosis 'maligniteit cervix').

Table 2. Use of Paclitaxel in Different Treatments

Treatment	12	13	103	113	803	1101	3101	3103	3201	3301	3302	4302	9101	Blank
Absolute	1	1	3	1	2	2	19	1	3	1	1	2	5	26
Relative	1,5%	1,5%	4,4%	1,5%	2,9%	2,9%	27,9%	1,5%	4,4%	1,5%	1,5%	2,9%	7,4%	38,2%

4 Conclusion

With this contribution we aimed at stimulating the development of new perspectives on care-flow mining. We investigated two new focuses with the use of the event log of the Department of Gynaecology at an academic hospital in the Netherlands. Firstly, we proposed the subdivision of the general care-flow and reviewed the patient evolution from the perspective of a specific department (radiotherapy in this case). Secondly, we further focussed on the use of Paclitaxel a drug that is commonly used for cancer chemotherapy. With this contribution we were able to confirm the usability of different focuses (i.e. beyond process discovery) on process mining techniques for retrieving useful information in a health-care setting.

References

1. Reijers, H., Russell, N., Geer, S., Krekels, G.: Workflow for healthcare: A methodology for realizing flexible medical treatment processes. In: Business Process Management Workshops, Springer (2010) 593–604
2. Van Dongen, B.: Real-life event logs: hospital log. doi:10.4121/d9769f3d-0ab0-4fb8-803b-0d1120ffcf54 (2011)
3. Weijters, A., van der Aalst, W., de Medeiros, A.: Process mining with the heuristics miner-algorithm. Technische Universiteit Eindhoven, Tech. Rep. WP **166** (2006)
4. Van der Aalst, W., Van Dongen, B., Gunther, C., Rozinat, A., Verbeek, H., Weijters, A.: Prom: the process mining toolkit. In: Proceedings of the BPM. (2009) 1–4
5. Van Der Aalst, W., Reijers, H., Song, M.: Discovering social networks from event logs. Computer Supported Cooperative Work (CSCW) **14**(6) (2005) 549–593
6. van der Aalst, W., De Beer, H., van Dongen, B.: Process mining and verification of properties: An approach based on temporal logic. On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE (2005) 130–147
7. National Institute for Health and Clinical Excellence: NICE Guidance TA108. http://www.nice.org.uk/nicemedia/live/11596/33590/33590.pdf (2007)
8. Society for Endocrinology: Chemotherapy schedules. http://erc.endocrinology-journals.org/cgi/data/ERC-09-0033/DC1/1 (2011)
9. National Cancer Institute: Cancer Drug Information: Paclitaxel. http://www.cancer.gov/cancertopics/druginfo/paclitaxel (2011)