Computers in Biology and Medicine 44 (2014) 88–96

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/cbm

Monitoring care processes in the gynecologic oncology department

Filip Caron [a,*], Jan Vanthienen [a], Kris Vanhaecht [b,c,d], Erik Van Limbergen [e], Jochen De Weerdt [a,f], Bart Baesens [a,g,h]

[a] Department of Decision Sciences and Information Management, KU Leuven, Naamsestraat 69, 3000 Leuven, Belgium
[b] Department of Public Health and Primary Care, KU Leuven, Kapucijnenvoer 35/7001, 3000 Leuven, Belgium
[c] Western Norway Research Network on Integrated Care, Helse Fonna, Haugesund, Norway
[d] Dutch Institute for Healthcare Improvement (CBO), Churchillaan 11, 3527 GV, Utrecht, The Netherlands
[e] Department of Radiation Oncology, University Hospital Gasthuisberg, KU Leuven, Herestraat 49/7003, 3000 Leuven, Belgium
[f] Information Systems School, Science and Engineering Faculty, Queensland University of Technology, 2 George Street, Brisbane QLD 4000, Australia
[g] Vlerick Leuven Gent Management School, Vlamingenstraat 38, 3000 Leuven, Belgium
[h] School of Management, University of Southampton, Highfield Southampton, SO17 1 BJ, United Kingdom

ARTICLE INFO

Article history:
Received 5 March 2013
Accepted 19 October 2013

Keywords:
Clinical pathways
Process mining
Health information systems
Healthcare quality
Medical informatics
Care processes

ABSTRACT

The care processes of healthcare providers are typically considered as human-centric, flexible, evolving, complex and multi-disciplinary. Consequently, acquiring an insight in the dynamics of these care processes can be an arduous task.

A novel event log based approach for extracting valuable medical and organizational information on past executions of the care processes is presented in this study. Care processes are analyzed with the help of a preferential set of process mining techniques in order to discover recurring patterns, analyze and characterize process variants and identify adverse medical events.

1. Introduction

Care processes or clinical pathways are defined as "complex interventions for mutual decision making and for the organization of care processes for a well-defined group of patients. They provide a detailed path that standardizes the coordination of roles and the sequencing of the activities in a multi-disciplinary team for a specific clinical problem or medical procedure" [1,2]. Medical informatics and decision making researchers have embraced the concept of clinical pathways and have used clinical pathways for configuring a wide spectrum of healthcare information systems, e.g. in [3–8]. The main objective of this approach has been the improvement of the provided care by bettering the risk-adjusted patient outcomes, promoting patient safety, increasing patient satisfaction and optimizing the use of resources [9,10]. In reality, the care processes of individual patients deviate from the standardized path in order to accommodate requirements dictated by the patient's specific medical conditions or the local conditions of the healthcare organization [11,12].

In this contribution we present a novel process mining based approach for acquiring insight in the real health care dynamics and the diagnosis-treatment cycles of individual patients. A preferential set of process mining techniques is applied on the data collected by the information systems that support the gynecologic oncology care processes at a major European academic hospital. The approach enables the analysis and monitoring of the provided care and the elicitation of tacit knowledge. Note that the contribution presents an open-minded analysis of the data without any knowledge on the designed pathway, local conditions or specific medical conditions of the patients.

The remainder of the paper is structured as follows: first, the preferential set and the data sources are discussed. Followed by the results of the gynecologic oncology case study, including both the analysis and representation of the general care process information and the identification of adverse events and healthcare risks. The discussion section provides an interpretation of the process mining results and discusses the limitations of this case study. The final section concludes the paper.

2. Methods

2.1. Subjects

The study has been executed on the care processes of 1143 patients of the gynecologic oncology department of a major

* Corresponding author. Tel.: +32 16 32 65 58; fax: +32 16 32 66 24.
E-mail addresses: Filip.Caron@kuleuven.be, Filip.Caron@econ.kuleuven.be (F. Caron).

0010-4825/$ - see front matter © 2013 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.compbiomed.2013.10.015

European academic hospital. All diagnosis and treatment activities that were performed for these patients during a period of 3 years, from January 2005 to January 2008 were collected. The patients were all diagnosed with a cancer pertaining either to the cervix, the vulva, the uterus or the ovaries. In total 236 treatment combinations are covered by this set of patients. As will be discussed model comprehensibility will increase when the diagnosis treatment cycles of the different patients are more similar. Therefore, different patient subsets were retrieved for examining specific elements in the gynecologic oncology department, i.e. the radiotherapy and chemotherapy cancer treatment.

Patients eligible for the radiotherapy subset underwent at least one diagnosis (i.e. consultations of different types) or treatment (i.e. teletherapy, hyperthermia therapy and brachytherapy) activity that was performed by the radiotherapy sub department. After filtering out the non-relevant patients we were left with a radiotherapy subset that contains 329 patients. Model development and validation were executed on this subset according to the common conformance checking approaches in process mining [13].

The chemotherapy subset contains the 42 oncology patients that received at least once a paclitaxel based chemotherapy. As this subset will be used for a quality assessment of the care processes for individual patients without the construction of a model, no validation set has been created.

2.2. Data sources

The study presented in this contribution is based on an event log supplied by a major European academic hospital [14], one of the largest academic hospitals in the Netherlands. Healthcare activities for the gynecologic oncology patients have been registered and documented by the hospital financial information systems.

A myriad of information has been recorded for each individual activity including the patient identifier, the activity type, the timestamp, the performer's department indicators and information on both the diagnosis and treatment type. Exactly 150,291 activities or an average of 131.5 activities per patient (spread: 1–1806 activities per patient) were recorded. All activity types (677 in total) are fine-grained and directly related to diagnostic and therapeutic activities of the gynecologic oncology pathways on which the core processes are based. We filtered out all administrative related activities, such as 'order rate' (translation of the activity 'ordertarief'). An example extract of the data set can be found in Table 1.

2.3. Proposal for analyzing care processes with process mining

Process mining techniques enable the translation of huge amounts of activity data, covering the interaction of patients with the hospital, into well-organized descriptions of the actual care processes [11]. In the context of care processes, we distinguish and propose a wide variety of analysis types that can assist in exploring and assessing care process models containing aggregated information on the provided care over a certain period of time, as well as in the evaluation of the care provided to individual patients. An overview of the most important analysis types can be found in Table 2.

Exploratory analysis techniques and tools provide the open-minded analyst with the support needed to acquire novel insights on the actual care processes. These techniques can be easily classified according to the main perspective of the exploratory analysis on the care process. In the context of care processes we distinguish four perspectives [15]: the functional, the process, the organizational and the data perspective.

- Functional perspective analyses deal with the existence (absence) or coexistence of certain process elements, e.g. specific activities.
- Process perspective analyses focus on the exact ordering of the activities. Typical analyses include the construction of an aggregated model for the care processes (i.e. care process discovery) and the discovery of care process variants with the accompanying gap analyses.
- Organizational perspective analyses center on the (human) resource aspect of a care process, including an analysis of the responsibilities, the authorization issues or the social networks.
- Case data perspective analyses revolve around all other types of information that have been recorded in the activity log. Therefore, these techniques are rather suited for clustering the care processes of individual patients and for other correlation analyses.

Advanced care processes analyses techniques focus on assessing the efficiency obtained in the care process, the quality of the provided care and the conformance of the provided care with medical guidelines. Efficiency analyses enable the identification of bottlenecks, e.g. activities that require too much time due to limited availability of specialized resources, in the organization of the healthcare providers. Additionally, they enable the comparison of process variants based on a variety of performance metrics, e.g. number of needed treatment cycles before disease free survival. Quality analyses deal with assessing specific properties of activity sequences for specific patients and the analysis of uncovered adverse events. Conformance analysis compares the provided care with the designed and prescriptive process models, to come to general conclusions transcending the level of the individual patient.

2.4. Overview of applied techniques

In this section we discuss the preferential set of process mining techniques in the context of care processes. Table 3 links the individual techniques to the different analysis types that were

Table 1
Activity log extract.

Patient	Event type	Treat.	Dep.	Diagnosis	Time
155	Follow-up polyclinic consultation	61	SGNA	Gyn. tumors	1-Jan-05
156	Cytological examination vagina	61	LVPT	Gyn. tumors	1-Jan-05
156	Histological examination	61	LVPT	Gyn. tumors	1-Jan-05
275	Teletherapy	13	RATH	Gyn. tumors	1-Jan-05
275	Follow-up polyclinic consultation	13	SGNA	Gyn. tumors	1-Jan-05
336	Potassium flame photometry	603	CHE2	Malign cervix	1-Jan-05
336	Differential count	603	HAEM	Malign cervix	1-Jan-05
336	Determination trombocyte level	603	HAEM	Malign cervix	1-Jan-05
336	Count of leukocytes	603	HAEM	Malign cervix	1-Jan-05
10	Count of leukocytes	113	HAEM	Malign cervix	4-Jan-05
10	Determination trombocyte level	113	HAEM	Malign cervix	4-Jan-05
72	Differential count	3101	HAEM	Malign ovary	16-Jan-05

Table 2
Process mining techniques for clinical pathway analysis.

Exploratory pathway analysis			
Functional analysis	***Process analysis***	***Organizational analysis***	***Case data analysis***
• Existence/absence of activities • Activity co-existence • Additional analyses	• Clinical pathway discovery • Process variant analysis • Recurring patterns • Additional analyses	• Social network analysis (teams, hand-overs, interactions) • Task allocation • Additional analyses	• Data-driven conditions • Correlations data and pathway structure • Additional analyses
Advanced pathway analysis			
Efficiency analysis		***Quality and conformance analysis***	
• Bottleneck analysis • Performance analysis and comparison of variants • Number and duration of diagnosis and treatment cycles • Additional analyses		• Rule-based pathway analysis • Conformance analysis • Analysis of adverse events • Root-cause analysis for variation • Additional analyses	

Table 3
Overview of the performed analyses with the applied techniques.

	Analysis sub type	**Analysis**	**Techniques/tools**	**Description**
Exploratory analysis	Functional analysis	Activity existence in a clinical pathway	Heuristics miner [13] and LTL-checker [14]	Presence of hyperthermia
	Process analysis	Clinical pathway discovery	Heuristics miner [13]	Radiotherapy workflow
		Clinical pathway variant analysis	Heuristics miner [13]	Radiotherapy workflow variant with hyperthermia
		Recurring patterns in individual patient care	Trace alignment [15]	Medical tests preceding the administration of paclitaxel
	Organizational analysis	Standard interaction and coordination between medical experts	Performance sequence diagram [16,17], Social network miner [18]	Interaction between radiotherapists and other medical specialists
	Case data analysis	Correlation between the activity log data and the activity sequence for individual patients	Custom-made rule patterns and process exploration	Correlation between (1) the presence of hyperthermia and the diagnosis codes, (2) the administration of paclitaxel and the diagnosis codes, (3) the administration of paclitaxel for ovarian cancer
Advanced analysis	Efficiency analysis	*Timestamps were to imprecise for efficiency analyses*		
	Quality and Conformance analysis	Rule-based pathway analysis	LTL-checker [14]	Compliance with treatment sequence
		Analysis of exceptional events	Custom-made rule patterns and process exploration	Absence of platinum agents for treating ovarian cancer with taxanes

previously discussed and that will be applied in the gynecologic oncology case study.

Heuristics miner is based on an activity dependency graph, i.e. sequences between activities, where evidence is collected to confirm or discard an activity sequence relationship [16]. The cutoff values determining the necessary level of evidence for confirming the existence of the activity sequence relationship, which enables the algorithm to deal with noise. On the other hand, highly infrequent behavior (such as adverse events) can be visualized by keeping the cutoff values low. In addition to effectively dealing with noise, experiments have proven that the use of the heuristics miner results in understandable process models even in the case of non-trivial constructs and low structured domains (such as care processes) [16].

Individual medical guidelines can often be matched to a particular rule structure, e.g. the required activity sequence structure, the activity deadline structure, etc. As a consequence generic business rule patterns can be defined for these recurring structures. At analysis time these generic patterns can be exactly configured based on the environmental elements, i.e. the activity type, the required originator role, etc. Ambiguity in the interpretation of the patterns is avoided by formally grounding them into linear temporal logic (LTL). LTL-checker provides the analyst with a tool to effectively assess the configured rule patterns against the event logs of a medical information system. The configured rule patterns are converted into a deterministic stream automaton which is afterwards used to check the event log [17]. While the LTL-checker provides an extensive rule pattern library in the context for functional and process analyses, the collection of rule patterns for analyzing and evaluating the organizational and case data perspective remains inadequate. In [18], we propose a rule pattern library of more than 60 individual and atomic patters (advanced rule patterns), which cover all the previously discussed perspectives. Together they provide the analyst with the most extensive set of rule patters, compared to tools such as the SCIFF checker [19]. Moreover, these techniques do not require any additional coding, as is the case in for instance [20]. Note that the rule pattern library remains extendible.

Trace alignment, which is based on the idea of biological sequence alignment, aims at organizing the traces in an event log in such a way that both common and exceptional behavior can be easily distinguished [21]. The plugin employs a two-step approach: (1) grouping similar traces in clusters and (2) visualize these clusters. The latter step is based on scoring matrices (e.g. substitution and indel), the generation of a guide tree with the agglomerative hierarchical clustering approach and a pruning a refinement based on alignment quality indicators.

Performance sequence diagram, this tool focuses on representing the transfer of work between data-element (e.g. departments or roles) instances over time [22,23]. Additionally, these representations enable the visualization of throughput times, the time in block or the average duration of the trace in a specific data element and the time in arrow or the average duration of the trace between different blocks. While this information can also be obtained through combinations of LTL-rules, the performance sequence diagram provides a more efficient and effective approach to obtain and visualize them.

The social network miner [24] is the only academic tool that provides support for generating and visualizing social networks based on the following five metrics; handover of work, subcontracting, working together, similar tasks and reassignment. Handover of work is based on causal dependency between activities, with options to consider only direct succession or take into account a causality fall factor. Subcontracting takes into account both the positive and negative evidence for the hypothesis that an activity (sequence) is performed by individual j in-between two activities executed by individual i. Working together focuses on how frequently certain individuals work together on the same case, while not taking into account the activity dependency. Different metrics have been implemented in this context. The similar tasks metric enables the analyst to define profiles for specific groups of individuals. Differences between the profiles can be measured using one of the following distance metrics: Euclidean distance, Pearson's correlation coefficient, similarity coefficient and the hamming distance. Finally, the reassignment metrics look at delegation of work between individuals.

Note that examining the full activity log and consequently the complete aggregated care process model does not result in comprehensible visuals. Therefore a filtering approach will be needed to obtain comprehensible and compact models. Filtering the activity log based on a specific data criterion, results in smaller event logs that contain more homogeneous behavior. In a care process analysis project filters could be placed on the diagnosis type, the treatment type, the performer type, the drug type or on the patient identifier. All techniques, mining algorithms and analysis tools that were used in this study are implemented in ProM, which is a generic open-source framework for process mining.

3. Results

This section first presents the results of the analyses performed on the radiotherapy subset, followed by the analysis results of the chemotherapy subset.

3.1. Radiotherapy in the gynecologic oncology care processes

3.1.1. Process discovery: Radiotherapy workflow

Fig. 1 depicts the retrieved care process model of the activities performed by the radiotherapy department. This process model consists of three clearly distinguishable blocks: (1) the medical consults, (2) the preparatory activities – i.e. the "simulation" activity and the "treatment time" calculation activity – and (3) the radiotherapy activities themselves: teletherapy, hyperthermia therapy and brachytherapy.

The heuristics miner [16] with standard parameterization has been used for the construction of this care process model. Models obtained with the heuristics miner typically represent general behavior, as it uses frequency thresholds for the selection of the discovered activity ordering relations that will be depicted. The numbers on the arcs (representing individual activity sequences) in the process model indicate respectively the frequency- and counterevidence-based level of certainty of the existence of that activity sequence relationship (between 0 and 1) and the number of times that sequence between the activities has taken place.

3.2. Compliance with standard activity sequence in radiotherapy

In this section we further investigate the common treatment sequence – teletherapy, hyperthermia therapy and brachytherapy – that has been suggested in the retrieved care process model.

Of the 215 care process instances containing at least one radiotherapy treatment, 190 instances (89%) start with teletherapy. Hyperthermia only occurs in a minority, 22 cases or approximately 10%, of the care process instances. Extensive testing of these 22 cases revealed that in approximately 27% the hyperthermia therapy is followed by an additional teletherapy treatment. In general, we conclude that teletherapy and hyperthermia are coupled and that teletherapy precedes hyperthermia. In 144 pathway instances the brachytherapy occurs at the end of the care process. In 83% of the cases including brachytherapy, the brachytherapy occurs directly after the teletherapy. Approximately 93% of the cases that include hyperthermia end with brachytherapy.

The adherence to this treatment sequence has been assessed with the LTL-checker tool [17]. LTL-checker allows for testing the activity log against (un)expected/(un)desirable process properties that are specified in temporal logic.

3.3. In-depth analysis of variants with hyperthermia therapy

The heuristics net in Fig. 1 shows that in a minority of the care process instances the hyperthermia therapy must be included, i.e. 26 times spread over 22 patients. In this section we further analyze the process variant that includes hyperthermia therapy.

3.3.1. Local care process discovery and social network analysis

Fig. 2 presents an enlargement of the local surroundings of the hyperthermia therapy (based on heuristics miner) in the care process model based on the complete activity log. This enables the exploration of departmental collaborations and the resulting needs for communication.

Typically, hyperthermia is preceded by an activity performed at the radiotherapy group; in exactly half of the cases the preceding activity is 'teletherapy' and in the other half 'treatment time'. Whereas the outflow may seem diverse, we can conclude that in the majority of the cases the patients are kept for further observation, i.e. day care or (clinical) hospitalization. In the social network related to this particular process part (Fig. 3), this collaboration or work-handover [24] is presented by the arc from the radiotherapy (hyperthermia) to the nursing ward.

3.3.2. Correlation between hyperthermia and diagnosis

No hard evidence was found for hyperthermia as a crucial treatment for any medical condition or (hospital specific internal) diagnosis code combination (Table 4). The code combinations have been abstracted. The highest (relative) occurrence frequency can be found for the combinations {e,b} and {a,b,c}, respectively 25% and 21%.

3.4. Investigating paclitaxel-based chemotherapy

3.4.1. Correlation between the administration of paclitaxel and diagnosis codes

Table 5 provides an overview of the diagnosis types for which we found patients who have been administered paclitaxel. The absolute and relative frequency is provided for each individual

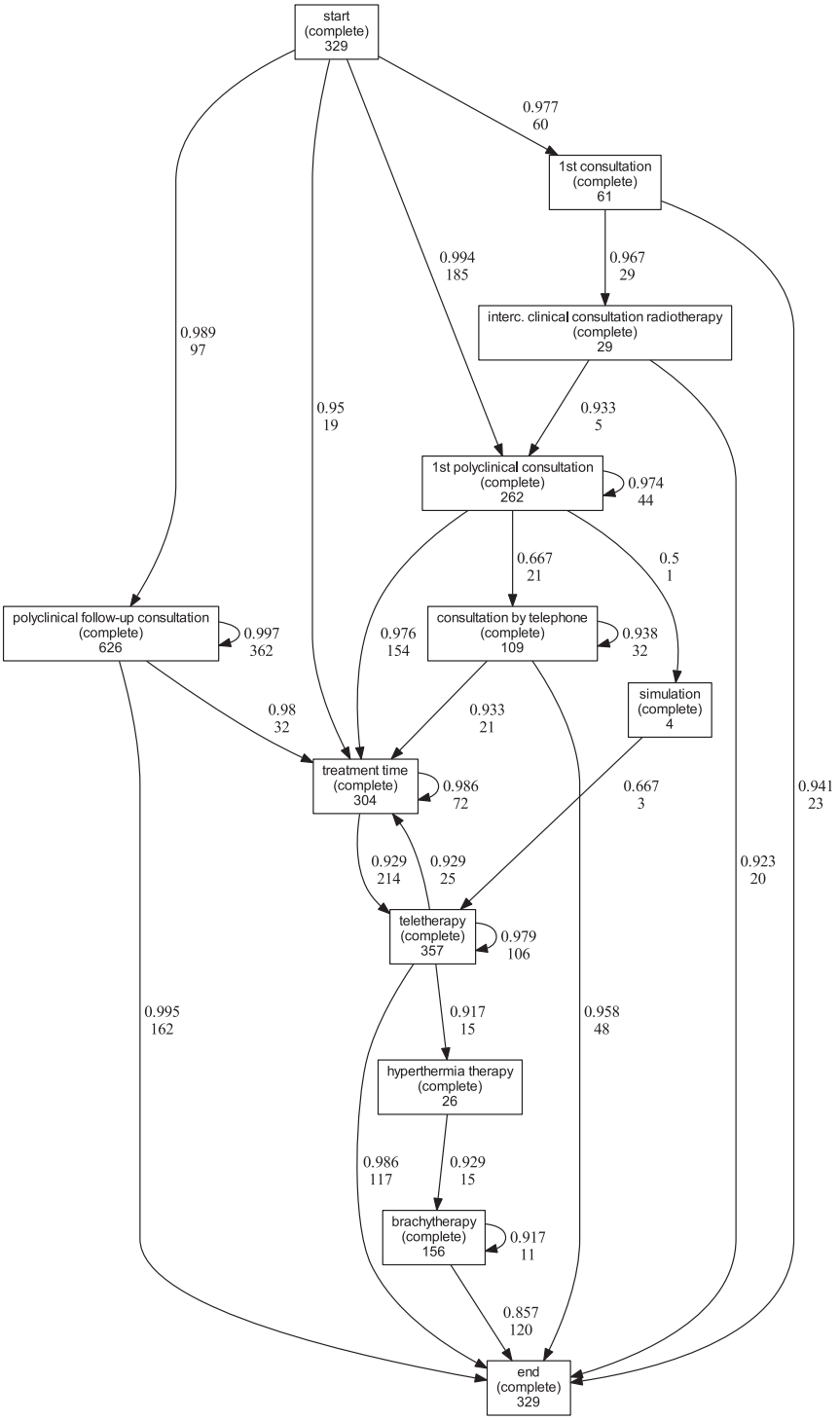


Fig. 1. Retrieved Clinical Pathway for Radiotherapy (Heuristics Miner).

diagnosis description and for the aggregated level. Note that different descriptions for highly similar diagnoses exist.

3.5. *Identifying the preceding medical tests*

Fig. 4 provides the trace alignment [21] of the care process parts directly preceding the administration of paclitaxel and that are performed by the 'general clinical chemistry lab' on patients diagnosed with 'malign neoplasm ovary'. Trace alignment is a technique based on the sequence alignment in bioinformatics and allows for the identification of regions of similarity. If needed gaps are inserted between the activities of a patient activity sequence so that similar activities are aligned in specific columns. We first grouped the patient activity sequences that are identical in order not to overburden the visual representation.

Each row in the left-hand side matrix represents a set of patients with identical laboratory test sequences, with the patient identifier in front denoting an example patient. Five medical tests typically recur in the following sequence: the determination of hemoglobin levels, creatine concentrations, sodium concentrations and thrombocyte

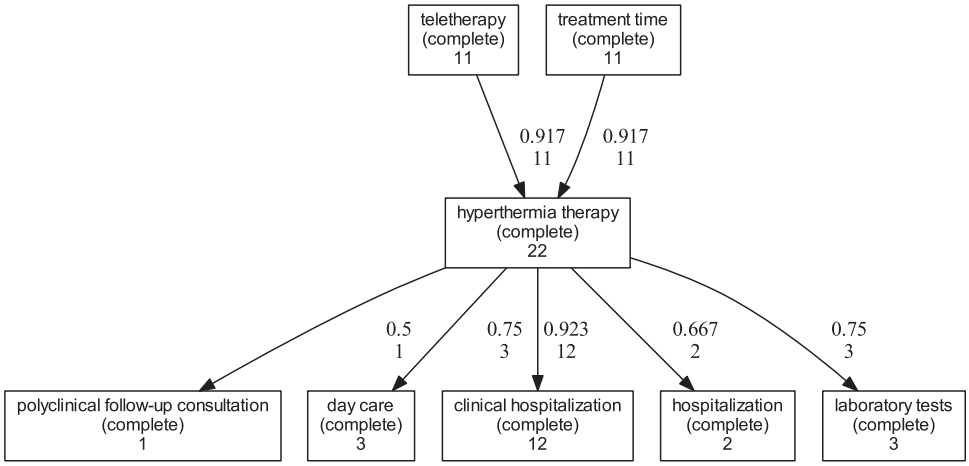


Fig. 2. Local heuristics net for hyperthermia therapy (only in- and out-flow for hyperthermia).

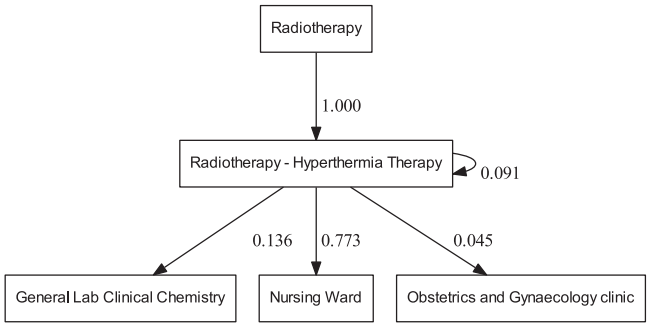


Fig. 3. Local social network around hyperthermia.

Table 4
Overview of the relation between hyperthermia therapy and diagnosis codes.

Diagnosis code combinations	Diagnosis description	Frequency (%)
{M13,106,822}	Malign cervix or Plaveiselcelca cervix or relapse malign cervix uteri	10 (21%)
{106,822}	Malign neoplasm cervix uteri	6 (10%)
{M13,106}	Malign cervix	3 (5%)
{106}	Malign melanoma vulva	1 (1,5%)
{M11,106}	Malign vulva	1 (9%)
{M12,106}	Plaveiselcelca vagina	1 (25%)

level as well as differential counting. Both the potentiometric titration of potassium and the count of leukocytes tend to be present as well. The information score right-hand side of Fig. 4 results in a consensus sequence with the histogram indicating the confidence in the position (and occurrence) of each element in the sequence.

3.6. Paclitaxel treatment for ovarian cancer

In this section, we investigate the use of paclitaxel for ovarian cancer over the different stages, i.e. drug-diagnosis multi-perspective. The ovarian cancer disease stage has a significant influence on the recommended chemotherapy [25].

3.6.1. Ovarian cancer stage Ia

In [26] it is concluded that patients diagnosed with ovarian cancer stage Ia do not require postoperative chemotherapy based on the active ingredient paclitaxel. Accordingly, the three patients

Table 5
Overview of the relation between paclitaxel administration and diagnosis codes.

Diagnosis description	Frequency (%)
Malign ovarian	**29 (37%)**
Malign neoplasm ovarian	9 (82%)
Malign ovarian	5 (83%)
Malign neoplasm adnexa uteri	4 (80%)
Malign neoplasm ovarian (nno)	1 (9%)
Ovarian cancer	1 (100%)
Serous adenoca: ovarian st IIIc	6 (18%)
Mucinous adenoca, ovarian st IIIc	1 (25%)
Relapse borderline malign ovarian	1 (100%)
Serous adenoca: ovarian st IV	1 (17%)
Malign tuba	**5 (9%)**
Malign ovarian|tuba	4 (7%)
Plaveiselcelca cervix st IVb	1 (100%)
Malign endometrium/corpus uteri	**4 (5%)**
Adenoca: corpus uteri st IIIa	1 (50%)
Adenoca: corpus uteri st IVb	1 (14%)
Relapse corpus malign uteri	1 (7%)
Malign endometrium	1 (2%)
Malign cervix	**3 (2%)**
Malign neoplasm cervix uteri	2 (25%)
Malign cervix	1 (<1%)
Gynecological tumors (not further specified)	**1 (2%)**

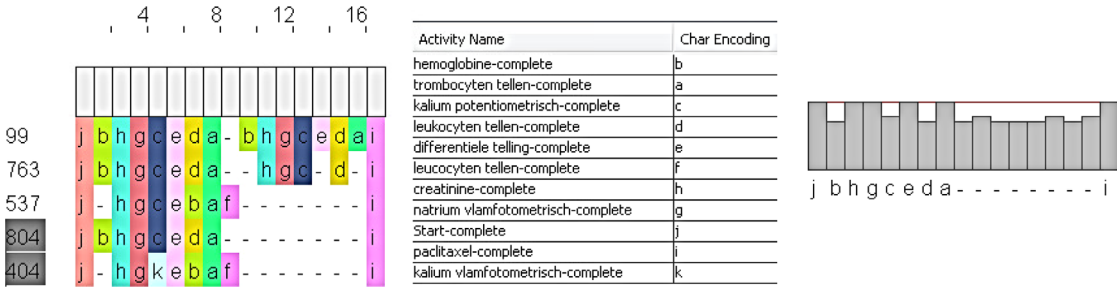
(patient 1049, 1067 and 1139) diagnosed with stage Ia ovarian cancer did not receive this type of chemotherapy. However, we only found activities performed by the operating room in the care path of patient 1049.

3.6.2. Ovarian cancer stage Ic and II

Further postoperative chemotherapy is recommended for patients with ovarian cancer stage Ic or stage II [27]. However, out of the five patients with this medical condition, only three have undergone surgery (patient 819, 987, 1020). But none of them received a postoperative paclitaxel treatment of at least three cycles, as described in [28].

3.6.3. Ovarian cancer stage III

In their literature review Herzog and Herrin [25] list three standard (after surgery) chemotherapy courses for stage III ovarian cancer. Each of these courses covers six 21-day-cycle treatments based on a combination of paclitaxel with carboplatin or cisplatin. While the majority of the paclitaxel administration schedules start with three 21-day-cycles, only the schedule for patient 829 comes



Activity Name	Char Encoding
hemoglobine-complete	b
trombocyten tellen-complete	a
kalium potentiometrisch-complete	c
leukocyten tellen-complete	d
differentiele telling-complete	e
leucocyten tellen-complete	f
creatinine-complete	h
natrium vlamfotometrisch-complete	g
Start-complete	j
paclitaxel-complete	i
kalium vlamfotometrisch-complete	k

Fig. 4. Local trace alignment for pathways including paclitaxel administration.

close to the six 21-day cycle treatment. Based on the activity log we were unable to conclude that any patient received the combination of paclitaxel and carboplatin or cisplatin.

4. Discussion

This section discusses the wide variety of insights into the actual provided healthcare that can be directly derived from the presented study results.

4.1. Radiotherapy in the gynecologic oncology care processes

Based on the retrieved care process model for the radiotherapy department we were able to uncover different types of rather exceptional behavior. First, this care process model indicates that a high level of the patients in this workflow never receive a radiotherapy treatment, i.e. transitions between a consultation related activity and the end state.

Second, under extremely exceptional circumstances, only four times according to the event log, a simulation activity has been explicitly performed. As a simulation is crucial and the non-execution of this type of activity could result in major quality risks, we assume that the registration of this activity has not been done systematically or that it is contained in the "treatment time" determination activity.

Third, while we established a clear sequence between the treatment activities, several patients were treated in a different way. For example, multiple patients diagnosed with cancer pertaining the cervix that received radiotherapy treatment (indicated by the presence of teletherapy), did not receive brachytherapy (e.g. patients 447, 451 and 457). However, according to Gerbaulet et al. brachytherapy is a standard of care for cervical cancer [29]. Many plausible explanations may exist, such as the patient's decision to stop the cancer treatment or the performance of the brachytherapy at a different hospital.

4.2. Investigating paclitaxel-based chemotherapy

Three major deviations from medical guidelines were uncovered during the analysis of the administration of paclitaxel-based chemotherapy: the versatile use of paclitaxel, the lack of combining platinum agents with paclitaxel and the lack of a postoperative paclitaxel-based chemotherapy (and of the surgery itself).

First, the study results indicate that the medical experts have been administering paclitaxel to treat cancers and tumors at the ovaries, the uterus, the cervix and the endometrium. Moreover, the event log contains a pathway instance for which the patient was diagnosed with the unspecific "gynecological tumors".

According to the label for "Bristol Myers Squibb Taxol" with active ingredient paclitaxel, published by the Food and Drug Administration on May 2nd, 2011 at [30], Taxol could be used to treat ovarian, breast and lung cancer as well as Kaposi's sarcoma. The label does not mention the usage of Taxol for the uterus, the cervix and the endometrium. This deviation could indicate an advanced insight in the effects of paclitaxel by the Dutch specialists.

Second, gynecologic oncology research has extensively studied the optimal composition of chemotherapy regimens for the treatment of ovarian cancer, both for the first- and second-line treatment. While there has been no general agreement on the optimal composition of the regimens, the beneficial effects (e.g. overall survival or progression free survival) of platinum agents (i.e. cisplatin or carboplatin) in the treatment of ovarian cancer have been demonstrated, e.g. [28,31]. Additionally, the NICE TA91 report considers the combination of a platinum-based compound with paclitaxel as a cost effective use of the NHS resources for second-line treatment of women with (partially) platinum-sensitive advanced ovarian cancer [32]. Furthermore, the "Assessment Group" analysis in [32] estimates that the use of a paclitaxel-platinum treatment could lead to 1.56 QALY, compared to 0.79 QALY for single-agent paclitaxel [33].

We were unable to find any events related to the administration of a platinum agent in the context of a chemotherapy treatment for ovarian cancer. Neither did the event log contain the trade names "Platinol" (for cisplatin) and/or "Paraplatin" (for carboplatin). Single-agent paclitaxel has been proposed by the NICE TA91 as a cost effective treatment for women with advanced ovarian cancer that are allergic to platinum [32]. However, the same technical appraisal states that only a small portion of the ovarian cancer patients in the United Kingdom is allergic to platinum. Another option for platinum allergic patients would be the use of doxorubicin hydrochloride encapsulated in pegylated liposome. Three individual care paths (for patient 209, 796 and 810) have been identified for this doxorubicin hydrochloride encapsulated in pegylated liposome treatment option.

Third, the in-depth investigation of the activity sequences of the patients with ovarian cancer that received a paclitaxel based chemotherapy uncovered two important deviations. The activity log did not contain a surgery for 2 out of the 3 patients that were diagnosed with ovarian cancer stage Ia, while this is adviced in [26]. In addition, while patients diagnosed with ovarian cancer stage Ic and stage II have an estimated disease recurrence rate between 25 and 40% and therefore require postoperative paclitaxel chemotherapy, we did not find any patient that received this type chemotherapy.

Note that the different descriptions of the diagnoses in Table 5 may indicate the absence of a systematic coding of the diagnosis, such as the International Classification of Diseases (ICD).

4.3. Limitations

The limitations of the current study can be grouped in three categories: the limitations of the constructed event log, the local hospital conditions and the lack of prescriptive clinical pathways.

The event log inherently suffered from the following limitations: a limited accuracy of the timestamps making performance analyses irrelevant and coarsely recorded additional event information (e.g. the originator information consisted of high-level departmental information). Moreover, we cannot guarantee that the log with a limited time frame covers all the activities of all the patients. In combination with the lack of additional patient information this results, for example, in the inability to determine whether patients were under first-line or second-line chemotherapy. The data is retrieved from a financial information system and consequently might not cover all the (non-invoiced) activities in the actual process. Second, the event log that was used in this care process analysis study was supplied by a university hospital, which is an important local condition as the log may contain a disproportionate amount of special and/or high risk cases. Finally, due to the absence of a designed prescriptive clinical pathway model we were unable to compare it directly with the actual care processes.

These limitations, however, do not affect the potential of applying process mining techniques for the continuous quality and performance monitoring of the care processes.

5. Conclusion

While clinical pathway design and implementation efforts are directed towards safeguarding both healthcare quality and performance, important deviations in the actual care processes from the medical guidelines may be required. The opportunities of process mining in this setting are twofold. First, the analysis techniques allow for the acquisition of a concise insight into the care processes and the resulting reorganization support. For example, the addition of platinum agents to the paclitaxel treatment might improve the patient outcomes. Second, these analysis techniques enable to either confirm compliance with the medical guidelines or to filter-out and further analyze the care process instances that violate the guidelines.

Summary

Background

Clinical pathways have been recognized as an effective technique to improve the risk-adjusted patient outcomes, to promote patient safety, to increase patient satisfaction and to optimize resource use. Patient specific medical conditions and/or the local conditions at the healthcare provider often require the actual care processes to deviate from the standard interventions as specified in the clinical pathways.

Results

The analysis of the diagnosis-treatment activity sequences resulted in the following outcomes: (1) the retrieval of the actual care processes, (2) irregularities in the care processes such as deviations from the common treatment sequence, (3) the use of paclitaxel for diagnoses that were not mentioned in the 'Taxol' label, (4) the absence of the paclitaxel-platinum combination for an effective chemotherapy and (5) activity sequences that did not contain the required operations. Note that we had no overview of a multitude of related medical factors.

Conclusions

With this study we demonstrate the applicability of process mining for monitoring the quality of healthcare provided in the context of specific type of care processes. These techniques provide valuable insight and assist with identifying potentially adverse medical events.

Conflicts of interest statement

None declared

References

[1] K. Vanhaecht, M. Panella, R.T. Van Zelm, W. Sermeus, What about care pathways? in: J.E. Ellershaw, S. Wilkinson (Eds.), Care of the Dying, second ed.,Oxford University Press, Oxford, 2010.

[2] H. Campbell, R. Hotchkiss, N. Bradshaw, et al., Integrated care pathways, Br. Med. J. 316 (1998) 133–137.

[3] R. Lenz, M. Reichert, IT support for healthcare processes—premises, challenges and perspectives, Data Knowl. Eng. 61 (2007) 39–58.

[4] A.M. Ryhämen, S. Rankinen, K. Tulus, et al., Internet-based patient pathway as an educational tool for breast cancer patients, Int. J. Med. Informatics 81 (2012) 270–278.

[5] R.W. Miller, A.G. Lee, J.S. Schiffman, et al., A practice pathway for the initial diagnostic evaluation of isolated sixth cranial nerve palsies, Med. Decis. Making 19 (1999) 42–48.

[6] K.M. Unerti, K.B. Johnson, N.M. Lorenzi, Health information exchange technology on the front lines of healthcare: workflow factors and patterns of use, J. Am. Med. Inf. Assoc. 19 (2012) 392–400.

[7] P.F.M. Stalmeier, Adherence and decision aids: a model and a narrative review, Med. Decis. Making 31 (2011) 121–129.

[8] K.M. Unerit, M.B. Weinger, K.B. Johnson, et al., Describing and modeling workflow and information flow in chronic disease care, J. Am. Med. Inf. Assoc. 16 (2009) 826–836.

[9] K. Vanhaecht, W. Sermeus, R.T. Van Zelm, et al., Care pathways are defined as complex interventions, BMC Med. 8 (2010) 31.

[10] T. Rotter, L. Kinsman, E. James, et al., Clinical pathways: effects on professional practice, patient outcomes, length of stay and hospital costs, Cochrane Database Syst. Rev. (2010) 3.

[11] Caron F., Vanthienen J., Deweerdt J., et al. Advanced care-flow mining and analysis, in: Daniel F et al., (Eds.), Volume 99 of LNBIP. Proceedings of the BPM 2011 Workshops, Part I; 28 August-2 September 2011, Clermont-Ferrand: Springer Verlag Berlin Heidelberg, 2012:167–168.

[12] K. Zheng, H.M. Haftel, R.B. Hirschl, et al., Quantifying the impact of health IT implementations on clinical workflow: a new methodological perspective, J. Am. Med. Inf. Assoc. 17 (2010) 454–461.

[13] A. Rozinat, van der Aalst WMP. Conformance checking of processes based on monitoring real behavior, Inf. Syst. 33 (2008) 64–95.

[14] Event Log Clinical Pathways. 2011. http://dx.doi.org/10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54 (accessed 27 Oct 2012).

[15] B. Curtis, M.I. Kellner, J. Over, Process modeling, Commun. ACM 35 (1992) 75–90.

[16] Weijters A., van der Aalst W.M.P., Alves de Medeiros AK. Process Mining with Heuristics Miner Algorithm, BETA Working Paper Series, WP 166. Eindhoven: Eindhoven University of Technology, 2006.

[17] van der Aalst W.M.P., De Beer H., van Dongen B.F. Process mining and verification of temporal properties: an approach based on temporal logic, in: Proceedings of On the Move to Meaningful Internet Systems 2005: CoopIS, DOA, and ODBASE; 31 October–4 November 2005, Agia Napan Cyprus, 2005:130–147.

[18] F. Caron, J. Vanthienen, B. Baesens, Comprehensive rule-based compliance checking and risk management with process mining decision support systems, Decis. Support Syst. 54 (2013) 1357–1369.

[19] M. Montali, Declarative Process Mining, Springer-Verlag, Berlin-Heidelberg, Germany, 2010.

[20] Robinson W.N., van der Aalst WMP. Monitoring Software requirements using instrumented code, in: Proceedings of the 35th Annual Hawai IEEE International Conference on Systems Sciences, IEEE Computer Society, 2002: 276–276.

[21] R. Bose, van der Aalst WMP. Process diagnostics using trace alignment: opportunities, issues and challenges, Inf. Syst. 37 (2012) 117–141.

[22] van der Aalst W.M.P., van Dongen B.F., Günther C., Rozinat A., et al. ProM: The process mining toolkit, in: Proceedings of the Seventh International Conference on Business Process Management; 8–9 September 2009, Ulm, 2009:1–4.

[23] W.M.P. van der Aalst, Challenges in business process analysis, Enterp. Inf. Syst. 1 (2009) 27–42.

[24] W.M.P. van der Aalst, H.A. Reijers, M. Song, Discovering social networks from event logs, Computer Supported Cooperative Work 14 (2005) 549–593.

[25] Herzog T., Herrin V. Patient Information: Medical Treatment of Epithelial Ovarian Cancer. 2011. ⟨http://www.uptodate.com/contents/patient-information-medical-treatment-of-epithelial-ovarian-cancer?view=print⟩ (accessed 15 Jan 2012).

[26] S. Kumar, H. Mahdi, C. Bryant, et al., Clinical trails and progress with paclitaxel in ovarian cancer, Int. J. Women's Health 2 (2010) 411–427.

[27] F. Robert, S. Ozols, G. Thomas, et al., Epithelial ovarian cancer, in: W.J. Hoskins (Ed.), Principles and Practice of Gynecologic Oncology, Lippincott Williams & Wilkins, Philadelphia, 2000, pp. 981–1057.
[28] J. Bell, M. Brady, R. Young, et al., Randomized phase III trial of three versus six cycles of adjuvant carboplatin and paclitaxel in early stage epithelial ovarian carcinoma: a gynecologic oncology group study, Gynecol. Oncol. 102 (2006) 432–439.
[29] A. Gerbaulet, R. Pöttter, J.J. Mazeron, et al., The GEC ESTRO Handbook of Brachytherapy, European Society for Therapeutic Radiology and Oncology, Leuven, Belgium, 2002.
[30] Bristol-Myers Squibb and Federal Drugs Administration. Taxol medical label. 2011. ⟨http://www.accessdata.fda.gov/drugsatfda_docs/label/2011/020262s049lbl.pdf⟩ (accessed 15 Jan 2012).
[31] M. Huizing, L. van Warmerdam, H. Rosing, et al., Phase I and pharmacologic study of the combination paclitaxel and carboplatin as firstline chemotherapy in stage III and IV ovarian cancer, J. Clin. Oncol. 15 (1997) 1953–1964.
[32] National Institute for Health and Clinical Excellence. Technology Appraisal 91: Ovarian Cancer: Paclitaxel, Pegylated Liposomal Doxorubicin Hydrochloride and Topotecan for Second-line or Subsequent Treatment of Advanced Ovarian Cancer. 2008. ⟨http://www.guidance.nice.org.uk/TA91/PublicInfo/pdf/English⟩ (accessed 15 Jan 2012).
[33] W. McGuire, A. Neugut, S. Arikian, et al., Analysis of the cost-effectiveness of paclitaxel as alternative combination therapy for advanced ovarian cancer, J. Clin. Oncol. 15 (1997) 640–645.

Filip Caron is a post-doctoral researcher at the Department of Decision Sciences and Information Management, Faculty of Economics and Business, KU Leuven, Belgium. His research interests include business process modeling and mining, care processes and decision support systems. Filip presented his findings at influential international conferences and has published them in renown international academic journals (e.g. Decision Support Systems, Expert Systems with Applications, Enterprise Information Systems and Computers in Industry). He is the winner of the 2011 Business Process Intelligence Challenge on care process data (category most complete analysis).

Jan Vanthienen is full professor of information systems at KU Leuven, Department of Decision Sciences and Information Management, Information Systems Group, where he is teaching and researching on business intelligence, analytics, business rules & processes, systems analysis, business information systems and information management. He has published more than 150 full papers in reviewed international journals and conference proceedings. He is a founding member of the Leuven Institute for Research in Information Systems (LIRIS), and a member of the ACM and the IEEE Computer Society. He was chairholder of the PricewaterhouseCoopers Chair on E - Business at KU Leuven, co-chairholder of the Microsoft Research Chair on Intelligent Environments. He received the Belgian Francqui Chair 2009 at FUNDP and an IBM Faculty Award in 2011. He is co-founder and president-elect of the Benelux Association for Information Systems (BENAIS). He is al so member of the IEEE task force on process mining, and co-author of the Business Process Mining Manifesto. He is actively involved in the request for proposal (RFP) for a Decision Modeling & Notation standard (DMN) at OMG (Object Management Group). This standard will be designed to complement the Business Process Modeling & Notation (BPMN) standard, in order to integrate and distinguish business processes and business decisions.

Kris Vanhaecht received his PhD in Public Health at Leuven University. He nowadays teaches quality in healthcare within the Department of Public Health and Primary Care and is policy advisor at the University Hospitals Leuven. He works part time as fellow at the Dutch Institute for Healthcare Improvement, CBO in Utrecht, is researcher at the Western Norway Research Network on Integrated Care, Helse Fonna and is the secretary general of the European Pathway Association.

Erik Van Limbergen MD, Ph.D., is senior Radiation Oncologist at the University Hospital Gasthuisberg in Leuven, Belgium. After obtaining a medical degree maxima cum laude at the Leuven University, he completed his training in radiation oncology at the University Hospitals in Leuven (1977–1981), at the Institut Gustave Roussy in Villejuif, and at Hôpital Henry Mondor, Créteil Paris, France (1981–1982). He obtained a certificate in Radiobiology at the University Paris XIII at Kremlin Bicêtre (1982). He obtained a Ph.D. with his thesis on: Breast Conserving Treatment for Breast Cancer: a decision making model based on qualitative and quantitative clinical parameters Since 1983 he is staff member at the University Hospitals in Leuven working in the field of breast and gynaecological cancers and soft tissue sarcomas and brachytherapy. Prof Van Limbergen is an active member of the European Society for Therapeutic Radiology and Oncology (ESTRO) and past-chair of the GEC-ESTRO Brachytherapy Committee (2006). He is director of the GEC-ESTRO Teaching Course of Modern Brachytherapy Techniques since 2000. He was chair of the breast programme in the first World Brachytherapy meeting in Boston 2008. He is an active member of the European Breast Cancer Network. He is treasurer of the EUropean Society of MAstology since 2007. He is author of more than 230 articles, published abstracts, and book chapters.

Jochen De Weerdt received a Master's degree in Business Economics—Information Systems Engineering from KU Leuven, Belgium. He is currently employed as a scientific researcher at the Department of Decision Sciences and Information Management at the KU Leuven. His research interests include data mining, process mining, and web intelligence.

Bart Baesens is an associate professor at KU Leuven, Belgium, and a lecturer at the University of Southampton (United Kingdom). He has done extensive research on predictive analytics, data mining, customer relationship management, fraud detection, and credit risk management. His findings have been published in well-known international journals (e.g. Machine Learning, Management Science, IEEE Transactions on Neural Networks, IEEE Transactions on Knowledge and Data Engineering, IEEE Transactions on Evolutionary Computation and Journal of Machine Learning Research) and presented at international top conferences. He is also co-author of the book Credit Risk Management: Basic Concepts, published in 2008.