Process Mining in Oncology: a Literature Review

Angelina Prima Kurniati[1,2,*], Owen Johnson[1], David Hogg[1], Geoff Hall[3,4]
[1] School of Computing, University of Leeds, Leeds, UK
[2] School of Computing, Telkom University, Bandung, Indonesia
[3] Leeds Institute of Cancer and Pathology, St James's University Hospital, UK
[4] School of Medicine, University of Leeds, Leeds, UK
*Corresponding author, email: scapk@leeds.ac.uk

Abstract— Process mining, an emerging data analytics method, has been used effectively in various healthcare contexts including oncology, the study of cancer. Cancer is a complex disease with many complicated care requirements and there is an urgent need to improve the cost and clinical effectiveness of cancer care pathways. Process mining of the e-health records of cancer patients may play an important future role and this paper presents a literature review of process mining in oncology as a contribution to this research. The search produced 758 articles which were manually reviewed by title, abstract, and full paper text review to develop the original pool of papers. An in-depth ancestor search was used to gather additional articles from the references of the original pool. These steps resulted in 37 papers. Through a thematic review process, the papers were analysed and five themes emerged. These were: 1) process and data types; 2) research questions; 3) techniques, perspectives and tools; 4) methodologies; 5) limitations and future work. This review can: (i) highlight the potential value of process mining for improving cancer care processes (ii) provide a useful overview of the current work undertaken; (iii) help researchers to choose process mining algorithms, techniques, tools, methodologies and approaches; and (iv) identify research opportunities in this new field of study.

Keywords: process mining, data mining, oncology, cancer, clinical pathways

I. Introduction

Process mining is an emerging field which utilizes automatically generated event logs in information systems for data analytics. An event log contains time-stamped records of events which refer to activities undertaken for specific cases. Event logs may also store additional information about events such as resources linked to the activity, or data elements recorded with the event. Exploring the event log is a way to analyse the records of real activities within organisations. In process mining [1], a process is typically modelled as a directed graph structure, in which the nodes represent transitions (i.e. events that may occur, e.g. MRI scan, chemo treatment) and places (i.e. conditions). The directed arcs describe which places are pre- and/or post conditions for which transitions (signified by arrows). To allow for events to happen in parallel (e.g. analysis of blood test, and MRI scan) the current state of a process may be represented by multiple nodes. Those process models can follow the notations of Petri nets or UML activity diagrams and Business Process Modelling Notation (BPMN), widely used in process improvement literature and practice [2]. This structure can then be used for further analysis, such as detecting deviations, repairing or enhancing the model.

Process mining has been applied to many fields including healthcare where it has contributed to improving quality of care, patient safety, patient satisfaction and optimization of resources [3]. Healthcare is characterized by highly complex and extremely flexible patient care processes (care pathways) and many autonomous, independently developed information systems [3], [4]. Process mining offers the opportunity to develop deeper understanding of this complexity and, if applied to cancer patient records, may help improve cancer care pathways and outcomes for cancer patients.

Cancer is a group of diseases characterized by the uncontrolled growth and spread of abnormal cells. The World Cancer Report of 2014 [5] noted that "cancer is among the leading causes of morbidity and mortality worldwide" (page 16), with 8.2 million deaths in 2012 and approximately 14 million new cases. Oncologists have recognized for many years that cancer is not a single disease but a large group of diseases that can affect any part of the body [6]. There are at least 65 types of cancer [7], which make the choices in cancer care pathways particularly challenging.

Cancer care is highly multidisciplinary, involving specialists from medical, surgical and radiation oncology, pathology, radiology, rehabilitation medicine, and many other disciplines. Cancer can be reduced and controlled by implementing evidence-based strategies for cancer prevention, early detection of cancer and management of patients with cancer. The optimum treatment of cancer should be supported by developing knowledge about the causes of cancer and the most effective processes for preventing and managing the disease. Exploring the event logs related to cancer treatment using process mining is a promising way to support the understanding and improving the quality of cancer care processes.

No systematic literature review has been previously published in process mining in oncology. Therefore, this paper presents a literature review that aims to support discussion on this topic by: (i) highlighting the potential value of process mining for improving cancer care processes (ii) providing a useful overview of the current work undertaken; (iii) helping researchers to choose process mining algorithms, techniques, tools, methodologies and approaches; and (iv) identifying research opportunities in this new field of study.

II. Process Mining in Oncology

Process mining has been proven to be useful in healthcare processes for discovering process models from event logs [3], [8] for checking the conformance of these models to the logs [9], [10], and mapping networks of resources to processes [8], [11], [12]. The guiding principles and challenges were defined by an IEEE Task Force on Process Mining in the Process Mining Manifesto (2011) which aimed to "increase the maturity of process mining as a new tool to improve the (re)design, control and support of operational business processes" [13] (page 1). There are a range of process mining algorithms but most have problems when analysing event data from clinical workflows, either in failure to construct useful process models or in models which do not reflect reality, mainly because of incomplete and noisy input data [12]. Despite these flaws, process mining has potential in helping to understand everyday clinical workflows and their variations.

A previous literature review was completed by Eric Rojas et al [14] on process mining in healthcare. This identified 66 papers with associated case studies which were analysed according to 11 main aspects. In that paper, they found that the most process mining case studies in healthcare were in oncology. However, 6 of the 9 papers found refer to one oncology dataset which was made available for the Business Process Intelligence Challenge (BPIC) in 2011 [15]. There were therefore only four different oncology datasets at the time of the review. This finding suggests that there is still a great opportunity to use process mining in oncology with a wider range of datasets and clinical questions.

This paper reviews the current literature on process mining in oncology. It summarizes the existing publications by process and data types, research questions and methodology. The primary goal is to describe the current state of research on process mining in oncology and identify the future opportunities available.

To address the aims listed above, we posed the following questions in relation to each paper:

Q1: What specific cases have been investigated in oncology?

Q2: What research methods have previously been used in process mining implementation on oncology?

Q3: What are the results of the previous research in process mining in oncology?

Q4: What are the future research opportunities in this field?

III. Research Method

A. Search Process

The term 'process mining' was coined in 1998 by Wil van der Aalst [16]. The IEEE Task Force on Process Mining was then established in 2009 [13] to promote the topic of process mining. A search on 'process mining' alone would therefore not be optimal because it would not include work earlier than 1998. Similarly, there is no medical subheading (MeSH) term for process mining. To be effective the search therefore used other keywords which appeared frequently in the Rojas et al review [14] and highly cited process mining literature such as [12]. An exploratory review was performed to verify relevant keywords. Other phrases were also used and tested, such as workflow mining, process analysis and pathway mining, but had no effect on the search results. The final query used in this literature review was:

```
("process mining" OR "data mining" OR
"machine learning" OR "pathway analysis")
AND ("event log" OR "patient flow") AND
("oncology" OR "cancer").
```

The query was applied to PubMed, BMJ Open, Journals of Clinical Oncology, ACM DL and Google Scholar.

Checking the inclusion criteria was done through three steps (see Figure 1), which were: title-based checking, abstract-based checking, and full-text checking. In each step, an article was excluded from the pool if it was: (a) a duplication of other papers in the pool, or (b) not a peer-review conference paper or journal article, or (c) not relevant to process mining in oncology.

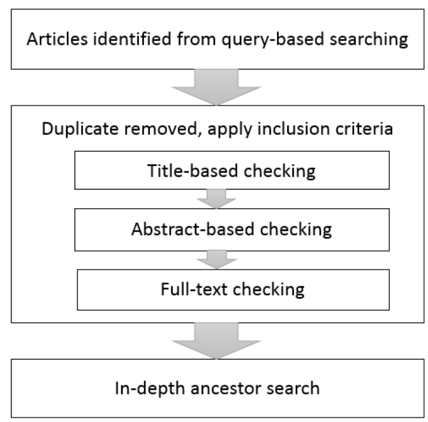

Figure 1. Review processes

In every step, a cautious filtering approach was implemented where a paper was excluded only if there were strong evidence that it was not relevant. If there were not enough information in the title or the abstract, respectively, the article was passed to the next step. The references of papers that passed the full-text checking were then investigated in an ancestor search to identify any possible additional relevant papers. The final set of papers was used for a thematic analysis.

B. Quality Assessment

To guarantee the quality of the search process a series of activities were undertaken. The initial query-based searching, analysis and evaluation of the articles were done manually by the first author. The query-based searching was done in July 2016. The Google Scholar search used Incognito Mode, a privacy feature which avoids bias in the search resulting from prior browsing history and automated recommendations that can confound search results [17], [18]. The co-authors supervised and verified all steps in the study, including the establishment of research strategies and review of the documents.

IV. RESULTS

The number of retrieved articles for each search engine is presented in Table I. All other search engines returned a subset of the 758 papers found by Google Scholar.

TABLE I. THE NUMBER OF RETRIEVED ARTICLES

Number	Search Engine	Number of articles
1	PubMed	3
2	BMJ Open	51
3	Journal of Clinical Oncology	28
4	ACM DL	74
5	Google Scholar	758

The title-based checking reduced the number of articles into 234, while abstract-based checking reduced them further into 97, and full text checking as the final checking step keep 33 papers in the pool. Table II shows the number of articles resulted from each step.

TABLE II. NUMBER OF ARTICLES IN EACH STEP

Step	Duplication	Not academic papers	Not relevant	Remaining
Title	12	23	489	234
Abstract	1	29	107	97
Text	0	23	41	33

The 37 papers were published between 2008 and 2016 (Table III). There were a number of papers which did discuss process mining before 2008 but none of these focused on oncology and therefore had been excluded. The availability of the BPIC 2011 dataset contributed to an increase in number of articles in 2013 onwards. The results demonstrate a small but steady increase in interest in process mining in oncology.

TABLE III. NUMBER OF ARTICLES PER YEAR

Year	Total	Analysis	
		BPIC 2011	Other data sources
2016	3	0	3
2015	9	6	3
2014	8	6	2
2013	10	8	2
2012	2	0	2
2011	2	2	0
2010	1	0	1
2009	1	1	0
2008	1	1	0
Total	37	24	13

A. Thematic Analysis

The first author read all 37 papers selected in the pool and identified common themes which were categorised and reviewed before being used for subsequent analysis. The five themes that emerged were: (1) process and data types; (2) research questions; (3) process mining perspectives, types and tools; (4) methodologies; (5) limitations and future work. The process and data types were further categorised for more detailed analysis. Research questions were similarly reviewed. Process mining perspectives, types and tools are the main theme in this review and an understanding of the gaps in the current literature will benefit other researchers. The papers were checked to see whether they follow one or more of the three perspectives suggested for process mining [1] (control-flow, performance and organizational perspectives) and one or more of the three types of process mining (discovery, conformance checking, and enhancement). The stated methodologies were analysed to identify whether the research followed a well-established methodology or was proposing a new methodology. The last theme examined limitations and future work including opportunities for future research in this field.

B. Process and data types

Processes analysed in the papers were coming from two general types, which were medical treatment processes and organizational processes [19]. To analyse these processes the researchers extracted data from administrative systems, clinical support systems, healthcare logistic systems, or automatically from medical devices.

The most commonly used dataset was from the Business Process Intelligence Challenge (BPIC) 2011 [15] which was used by 24 of the 37 papers (see Table III). This dataset is an anonymised event log from a Dutch Academic Hospital, containing 150,000 events for 1143 cases, each case being a patient attending the Gynaecology Department.

Figure 2 (below) shows an analysis based on cancer type. The most common being gynaecological cancer (24 papers) – all of which used the BPIC dataset. Of the other papers these covered breast cancer (4), colon, gastric and lung cancer (3 papers on each), rectal cancer (2), and bladder, cervical, head and neck, and skin cancer (1 paper on each). A complete list of data sources being used is available in Appendix 1.

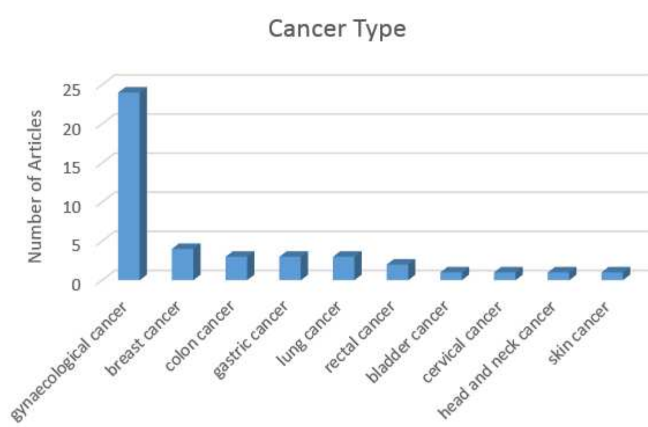

Figure 2. Number of articles based on cancer type

C. Research Questions

Most of the papers were working on answering research questions on the applicability of process mining in healthcare domains, specifically oncology. Some papers used oncology as a case study (or one of several case studies) to prove the concept being proposed, and others tried to solve problems in the dataset using different kinds of techniques.

Seven papers concern problems with heterogeneity in the data [11], [20]–[25]. Seven papers used declarative process mining and its enhancements [26]–[32]. Declarative process mining involves identifying possible execution paths for bridging event data reflecting the clinical reality and clinical guidelines describing best-practice. Two papers concerned the analysis of process anomalies and exceptions [33], [34], and some others concerned common pathways [35]–[38]. Problems and challenges of process characteristic issues and event log quality issues were well discussed in three papers [39]–[41].

Those research questions represented possible healthcare analysis supported by process mining [14], [42]:

- What happened: identifying the need to discover the process executed and its activities
- Why did it happen: understanding the activities and circumstances characterizing the situation/ action
- What will happen: identifying the circumstances of when or how a specific activity will take place
- What is the best that can happen: identifying possible steps towards specific improvements

D. Process Mining Techniques, Perspectives, and Tools

According to van der Aalst [43], there are three different perspectives in process mining: control-flow, performance, and organizational. Control-flow perspectives focus on the ordering of activities, performance perspectives test the performance of processes being analysed, and organizational perspectives analyse the working relations between parties in the process. There are also three main types of process mining: discovery, conformance, and enhancement [43]. Discovery builds a process model based on an event log. Conformance checks the conformance of the model discovered to the event log. Enhancement aims to extend or improve an existing process model using information about the actual process recorded in some event log.

All identified papers have applied at least one perspective and one type of process mining. All papers except one [39] discussed the control-flow perspective. Most of the papers (27 of the 37) discussed the performance perspective, but only 5 discussed the organizational perspective [3], [11], [23], [44], [45]. All papers, except two [35], [39], studied discovery from a control-flow perspective. Of these two, Ramezani et al. [35] presented an approach to facilitate creating and understanding formal compliance requirements by providing configurable templates using question trees and natural language. Rojas et al. [39] described several classes of process characteristics and data quality problems. Most of the papers (27 of the 37) applied conformance checking, but there were 8 papers applied enhancement.

In terms of tools being used, 24 papers using the ProM toolkit (www.promtools.org) [46]. The ProM toolkit is a de-facto standard in process mining research community and can be combined with other tools, such as GATE developer, WordNet, R, R Studio, and Java [25], Tilde, Alchemy and BUSL [47]. Other papers proposed their own tool [36], [38], [40], [48]–[53]. In case studies other than oncology, process mining has also been implemented using the DISCO commercial tool (www.fluxicon.com/disco), such as in [54]–[56]. Figure 3 illustrates how process mining perspectives, types and tools were discussed in the papers.

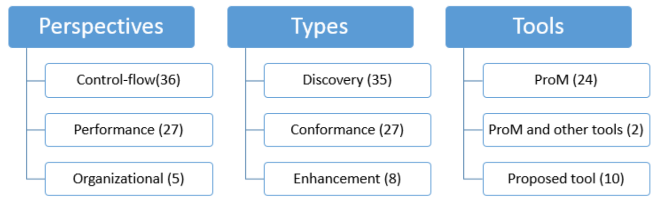

Figure 3. Process mining perspectives, types and tools

E. Methodologies

In the Process Mining Manifesto [13], van der Aalst et al suggested the L* life-cycle model describing a typical process mining project. This model consists of five stages: plan and justify, extract, create control-flow model and connect event log, create integrated process model, and operational support. Only one paper [57] clearly mentioned the L* lifecycle model as the methodology being used in the study. This implies limited awareness of a methodology proposed specifically for implementing process mining.

In general, the papers proposed new algorithms and/ or techniques in process mining implementation and used a common methodology to introduce the problem – they described a solution supported by the proposed algorithm/ techniques, and introduced a new prototype or plug-in implementing the proposed algorithm / techniques. These were then applied to their case studies. The eleven papers following this general methodology were [22], [27], [30], [32], [34], [35], [41], [47], [48], [58]–[60]. The methodology followed by the other ten papers [3], [11], [20], [29], [34], [37], [40], [44], [61], [62] was using available plug-ins and/ or functionalities in existing tools to solve the problem in their case study.

F. Limitations and future work

Limitations identified in the literatures can be classified as data, techniques, and team limitations. Data limitations were identified in ten papers [20], [22], [33], [36], [37], [44], [48], [57], [60] and were mostly related to limited access to the data, data quality problems, attributes not available in the data being extracted, or the dataset available in inappropriate level of details. Technique limitations were identified in 13 papers [3], [21], [28], [31], [35], [38]–[40], [48], [50], [52], [63], [64] specifically if the study undertaken was done by implementing plug-ins/ functionalities available in tools. Team limitations were identified in two papers [11], [23] because the authors realize that they need medical domain experts to be included in the research team.

Future work suggested in the papers followed the same classification as the limitations, namely data, technical, and team improvements. Some papers identified data improvements [28]–[30], [36]–[38], [48], [57], [64] in terms of improving data quality, dimensionality, and complexity. A larger number of important technical improvements were identified in 20 papers [3], [21]–[23], [29]–[31], [35], [39]–[41], [44], [48]–[52], [58], [60], [64]. These were to:

- Develop new mining techniques to obtain understandable, high-level information
- Implement clustering to view process variations better
- Consider medical correctness and relevance by a professional
- Elaborate networked graph visualizations and integrate with existing process discovery techniques
- Optimise the performance of the proposed technique
- Make the proposed framework suitable to be used in online settings

Team improvements identified in three papers [11], [23], [35] related to the multi-disciplinary nature of oncology treatments which required the research team to be expert in healthcare domain and technical domain as well.

V. Conclusion

Process mining is an emerging data analytics method which can benefit healthcare experts, allowing them to find process improvement opportunities. In oncology, this is especially important because oncologists and other experts from multidisciplinary fields can develop a better understanding of current care pathways in order to identify opportunities to improve the quality of cancer care.

This paper provide a useful bibliographic survey which gives an overview of background ideas, processes, methodologies, results and findings of current research in process mining in oncology. Some limitations and future work were identified and these should serve as a motivational guide to stimulate insights for future improvements in this field.

Acknowledgment

First author would like to thank the Indonesian Endowment Fund for Education (LPDP) for the sponsorship and all support given during this research.

References

[1] W. Van Der Aalst, Process mining: discovery, conformance and enhancement of business processes, 1st ed. Springer-Verlag Berlin Heidelberg, 2011.

[2] J. C. Recker, M. zur Muehlen, K. Siau, J. Erickson, and M. Indulska, “Measuring method complexity : UML versus BPMN,” Fac. Sci. Technol. Sch. Inf. Syst., 2009.

[3] R. S. Mans, M. H. Schonenberg, M. Song, W. M. P. van der Aalst, and P. J. M. Bakker, “Application of Process Mining in Healthcare – A Case Study in a Dutch Hospital,” Proc. BIOSTEC 2008, vol. 25, pp. 425–438, 2008.

[4] R. Lenz, T. Elstner, H. Siegele, and K. a Kuhn, “A practical approach to process support in health information systems.,” J. Am. Med. Inform. Assoc., vol. 9, no. 6, pp. 571–585, 2002.

[5] B. W. Stewart and C. P. Wild, “World Cancer Report 2014,” the International Agency for Research on Cancer, 2014. [Online]. Available: http://apps.who.int/bookorders/anglais/detart1.jsp?&codcch=275. [Accessed: 09-Aug-2016].

[6] American Cancer Society, “The History of Cancer,” 2011. [Online]. Available: www.cancer.net/patient/Advocacy and Policy/Treatment_Advances_Timeline.pdf. [Accessed: 09-Aug-2016].

[7] Cancer Research UK, “Your cancer type,” 2014. [Online]. Available: http://www.cancerresearchuk.org/about-cancer/type/. [Accessed: 09-Aug-2016].

[8] R. Mans, H. Reijers, M. van Genuchten, and D. Wismeijer, “Mining processes in dentistry,” Proc. 2nd ACM SIGHIT Symp. Int. Heal. informatics - *IHI ’12*, p. 379, 2012.

[9] J. B. van Osch Dekker, “Process Mining: Acquiring Objective Process Information for Healthcare Process Management with the CRISP-OM Framework I PREFACE,” Eindhoven University of Technology, 2008.

[10] K. Kirchner, N. Herzberg, A. Rogge-Solti, and M. Weske, “Embedding conformance checking in a process intelligence system in hospital environments,” Process Support Knowl. Represent. Heal. Care. Springer Berlin Heidelb., pp. 126–139, 2013.

[11] R. P. J. C. Bose and W. M. P. can der Aalst, “Analysis of Patient Treatment Procedures.,” Bus. Process Manag. Work., vol. 99, pp. 165–166, 2011.

[12] M. Lang, T. Bürkle, S. Laumann, and H.-U. Prokosch, “Process Mining for Clinical Workflows: Challenges and Current Limitations,” eHealth Beyond Horiz. - Get IT There (Proceedings 21st Int. Congr. Eur. Fed. Med. Informatics, MIE, pp. 229–234, 2008.

[13] W. van der Aalst, A. Adriansyah, A. K. A. de Medeiros, F. Arcieri, T. Baier, T. Blickle, J. C. Bose, P. van der Brand, R. Brandtjen, J. Buijs, A. Burattin, J. Carmona, M. Castellanos, J. Claes, and J. Cook, “Process Mining Manifesto,” Bus. Process Manag. Work., vol. 99, pp. 169–194, 2011.

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

[15] B. W. Boudewijn van Dongen, Diogo R. Ferreira, “Business Process Intelligence Challenge (BPIC),” IEEE Task Force on Process Mining, 2011. [Online]. Available: http://www.win.tue.nl/bpi/doku.php?id=2011:challenge. [Accessed: 22-Apr-2016].

[16] W. Van der Aalst, “The application of Petri nets to workflow management,” J. circuits, Syst. Comput., vol. 8, no. 01, pp. 21–66, 1998.

[17] G. Aggarwal and E. Bursztein, “An Analysis of Private Browsing Modes in Modern Browsers.,” USENIX Secur. Symp., pp. 79–94, 2010.

[18] H. Said, N. Al Mutawa, I. Al Awadhi, and M. Guimaraes, “Forensic analysis of private browsing artifacts,” 2011 Int. Conf. Innov. Inf. Technol. IIT 2011, pp. 197–202, 2011.

[19] U. Kaymak and R. Mans, “On process mining in health care,” IEEE Int. Conf. Syst. Man, Cybern., pp. 1859–1864, 2012.

[20] M. Song, C. Günther, and W. Van der Aalst, “Trace clustering in process mining,” Lect. Notes Bus. Inf. Process., 2008.

[21] B. Hompes and J. Buijs, “Discovering Deviating Cases and Process Variants Using Trace Clustering,” Proc. 27th Benelux Conf. Artif. Intell., pp. 5–6.

[22] C. Di Francescomarino and M. Dumas, “Clustering-Based Predictive Process Monitoring,” arXiv Prepr. arXiv1506.01428, 2015.

[23] F. Caron, J. Vanthienen, J. De Weerdt, B. Baesens, J. De Weerdt, and B. Baesens, “Beyond X-Raying a Care-Flow: Adopting Different Focuses on Care-Flow Mining,” Proc. First Int. Bus. Process Intell. Chall., pp. 1–11, 2011.

[24] L. Rosa, F. Maria, H. Nguyen, M. Dumas, M. La Rosa, F. Maria Maggi, and S. Suriadi, “Mining Business Process Deviance: A Quest for Accuracy,” Proc. [Lecture Notes Comput. Sci., vol. 8841, pp. 436–445, 2014.

[25] A. Deokar and J. Tao, “Semantics-based event log aggregation for process mining and analytics,” Inf. Syst. Front., vol. 17, pp. 1209–1226, 2015.

[26] M. Rovani, F. F. M. Maggi, M. De Leoni, and W. M. P. Van Der Aalst, “Declarative process mining in healthcare,” Expert Syst. Appl., vol. 42, no. 23, pp. 9236–9251, 2015.

[27] F. Maggi, R. Bose, and W. van der Aalst, “A knowledge-based integrated approach for discovering and repairing declare maps,” Int. Conf. Adv. Inf. Syst. Eng. Springer Berlin Heidelb., pp. 433–448, 2013.

[28] R. Bose, F. Maggi, and W. van der Aalst, “Enhancing declare maps

based on event correlations," Bus. Process Manag. Springer Berlin Heidelb., pp. 97–112, 2013.
[29] F. Maggi and M. Dumas, "Discovering data-aware declarative process models from event logs," Bus. Process Manag. Springer Berlin Heidelb., pp. 81–96, 2013.
[30] A. Burattin and M. Cimitile, "Online Discovery of Declarative Process Models from Event Streams," IEEE Trans. Serv. Comput., vol. 8, no. 6, pp. 833–846, 2015.
[31] A. Burattin, F. Maggi, and A. Sperduti, "Conformance checking based on multi-perspective declarative process models," arXiv Prepr. arXiv1503.04957, 2015.
[32] M. F.M. and F. Maggi, "Discovering metric temporal business constraints from event logs," Lect. Notes Bus. Inf. Process., vol. 194, pp. 261–275, 2014.
[33] J. Poelmans, G. Dedene, G. Verheyden, H. Van Der Mussele, S. Viaene, and E. Peters, "Combining business process and data discovery techniques for analyzing and improving integrated care pathways," *ICDM'10 Proc. 10th Ind. Conf. Adv. data Min. Appl.* Theor. Asp., pp. 505–517, 2010.
[34] S. J. J. Leemans, D. Fahland, and W. M. P. van der Aalst, "Discovering block-structured process models from event logs containing infrequent behaviour," Lect. Notes Bus. Inf. Process., vol. 171, pp. 66–78, 2014.
[35] E. Ramezani, D. Fahland, and W. van der Aalst, "Supporting domain experts to select and configure precise compliance rules," Int. Conf. Bus. Process Manag. Springer Int. Publ., pp. 498–512, 2013.
[36] D. Antonelli, E. Baralis, G. Bruno, S. Chiusano, N. A. Mahoto, and C. Petrigni, "Analysis of diagnostic pathways for colon cancer," Flex. Serv. Manuf. J., vol. 24, no. 4, pp. 379–399, 2012.
[37] F. Caron, J. Vanthienen, and B. Baesens, "Healthcare Analytics: Examining the Diagnosis–treatment Cycle," Procedia Technol., vol. 9, pp. 996–1004, 2013.
[38] A. Leontjeva and R. Conforti, "Complex Symbolic Sequence Encodings for Predictive Monitoring of Business Processes," Int. Conf. Bus. Process Manag. Springer Int. Publ., pp. 297–313, 2015.
[39] R. P. J. C. Bose, "Wanna Improve Process Mining Results ? It ' s High Time We Consider Data Quality Issues Seriously," Proc. 2013 IEEE Symp. Comput. Intell. Data Mining, CIDM 2013 - 2013 IEEE Symp. Ser. Comput. Intell. SSCI 2013, no. 1, pp. 127–134, 2013.
[40] J. De Weerdt, F. Caron, J. Vanthienen, and B. Baesens, "Getting a Grasp on Clinical Pathway Data: An Approach Based on Process Mining," Emerg. Trends Knowl. Discov. Data Min., pp. 22–35, 2012.
[41] X. Lu and R. Mans, "Conformance checking in healthcare based on partially ordered event data," Proc. 2014 IEEE Emerg. Technol. Fact. Autom., pp. 1–8, 2014.
[42] R. Mans, W. van der Aalst, and R. Vanwersch, Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes, 1st ed. Springer International Publishing, 2015.
[43] W. van der Aalst, "Data Science in Action," in Process Mining, 2nd ed., Springer-Verlag Berlin Heidelberg, 2016, pp. 3–23.
[44] F. Caron, J. Vanthienen, K. Vanhaecht, E. Van Limbergen, J. De Weerdt, and B. Baesens, "Monitoring care processes in the gynecologic oncology department," Comput. Biol. Med., vol. 44, pp. 88–96, 2014.
[45] T. Goila and A. Agrahari, "A Framework for Evidence Based Business Process Redesign," Available SSRN, 2014.
[46] B. Van Dongen, "The ProM framework: A new era in process mining tool support," Int. Conf. Appl. Theory Petri Nets. Springer Berlin Heidelb., pp. 444 – 454, 2005.
[47] E. Bellodi, F. Riguzzi, and E. Lamma, "Statistical Relational Learning for Workflow Mining}," Intell. Data Anal., vol. 20, no. 3, pp. 515–541, 2015.
[48] Z. Huang, W. Dong, L. Ji, C. Gan, X. Lu, and H. Duan, "Discovery of clinical pathway patterns from event logs using probabilistic topic models," J. Biomed. Inform., 2014.
[49] M. Räim, C. Di Ciccio, F. F. M. Maggi, M. Raeim, C. Di Ciccio, F. F. M. Maggi, M. Mecella, and J. Mendling, "Log-Based Understanding of Business Processes through Temporal Logic Query Checking," Move to Meaningful Internet Syst. Otm 2014 Conf., vol. 8841, no. October 2014, pp. 75–92, 2014.
[50] H. Syed and A. Das, "Temporal Needleman-Wunsch," IEEE Int. Conf. Data Sci. Adv. Anal. (DSAA), 2015. 36678 2015., pp. 1–9, 2015.
[51] F. Ju, H. K. Lee, R. U. Osarogiagbon, X. Yu, N. Faris, and J. Li, "Computer modeling of lung cancer diagnosis-to-treatment process.," Transl. lung cancer Res., vol. 4, no. 4, pp. 404–14, 2015.
[52] Z. Huang, X. Lu, H. Duan, and W. Fan, "Summarizing clinical pathways from event logs," J. Biomed. Inform., 2013.
[53] S. Hina, E. Atwell, and O. Johnson, "SnoMedTagger: A semantic tagger for medical narratives," Int. J. Comput. Linguist. Appl., vol. 4, no. 2, pp. 81–99, 2013.
[54] A. Partington, M. Wynn, and S. Suriadi, "Process mining for clinical processes: a comparative analysis of four Australian hospitals," ACM Trans. Manag. Inf. Syst., vol. 5, no. 4, p. 19, 2015.
[55] P. Delias, M. Doumpos, E. Grigoroudis, P. Manolitzas, and N. Matsatsinis, "Supporting healthcare management decisions via robust clustering of event logs," Knowledge-Based Syst., 2015.
[56] J. J. Boere, "An Analysis and Redesign of the ICU Weaning Process using Data Analysis and Process Mining," Maastricht University Medical Centre, 2013.
[57] M. Binder, W. Dorda, and G. Duftschmid, "On analyzing process compliance in skin cancer treatment: An experience report from the evidence-based medical compliance cluster (ebmc2)," Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), 2012.
[58] F. Maggi and C. Di Francescomarino, "Predictive monitoring of business processes," Int. Conf. Adv. Inf. Syst. Eng. Springer Int. Publ., pp. 457–472, 2014.
[59] U. Kaymak, R. Mans, T. Van De Steeg, and M. Dierks, "On Process Mining in Health Care," IEEE Int. Conf. Syst. Man, Cybern., pp. 1859–1864, 2012.
[60] J. Meier, A. Dietz, A. Boehm, and T. Neumuth, "Predicting treatment process steps from events," J. Biomed. Inform., vol. 53, pp. 308–319, 2015.
[61] F. Caron, J. Vanthienen, J. De Weerdt, and B. Baesens, "Advanced care-flow mining and analysis," in Lecture Notes in Business Information Processing, 2012, vol. 99 LNBIP, no. PART 1, pp. 167–168.
[62] B. Hompes, J. Buijs, and W. van der Aalst, "Detecting Change in Processes Using Comparative Trace Clustering," processmining.org. [Online]. Available: http://www.processmining.org/_media/blogs/pub2015/paper7.pdf. [Accessed: 02-Mar-2016].
[63] O. Johnson, P. S. Hall, and C. Hulme, "NETIMIS : Intelligent Simulation of Health Economics Outcomes using Big Data," Pharmacoeconomics, vol. 34, no. 2, pp. 107–114, 2015.
[64] Y. Sun, Y. Du, and M. Li, "A Repair of Workflow Models Based on Mirroring Matrices," Int. J. Parallel Program., pp. 1–20, 2016.
[65] F. Maggi, "Declarative Process Mining with the Declare Component of ProM.," BPM, 2013.

Appendix 1. Data Sources

Number	Reference	Summary					
		Year	process and data types	Patients	Cancer type	Origin	Description
1	[3]	2008	BPIC 2011	627	gynaecological cancer	Dutch	
2	[20]	2009	BPIC 2011	619	gynaecological cancer	Dutch	
3	[33]	2010	Inpatient data	148	breast cancer	Belgium	
4	[23]	2011	BPIC 2011	329	gynaecological cancer	Dutch	Sub process of Dept. of Radiotherapy
5	[11]	2011	BPIC 2011	1143	gynaecological cancer	Dutch	
6	[36]	2012	Supplied service data	134	colon cancer	Italy	From healthcare territorial agencies.
7	[57]	2012	Clinical and administrative data	389	skin cancer	Austria	Cutaneous Melanoma (CM) stage IV protocols
8	[39]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	Other datasets: Philips Healthcare, BPIC 2012, Catharina Hosp., CoSeLog
9	[65]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	
10	[40]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	
11	[28]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	
12	[29]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	
13	[35]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	
14	[34]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	Other datasets: BPIC 2012, logs of a building permit approval process in CoSeLog project.
15	[37]	2013	BPIC 2011	1143	gynaecological cancer	Dutch	
16	[52]	2013	Dataset from Zhejiang Huzhou Central hospital of China in two years	200	bronchial lung cancer, colon cancer, gastric cancer	China	bronchial lung cancer (48), colon cancer (52), gastric cancer (100), cerebral infarction (445)
17	[27]	2013	Event logs of patient treatment	289	bladder cancer	Dutch	
18	[44]	2014	BPIC 2011	329	gynaecological cancer	Dutch	oncology patients that received at least once a paclitaxel based chemotherapy
19	[45]	2014	BPIC 2011	1143	gynaecological cancer	Dutch	
20	[58]	2014	BPIC 2011	1143	gynaecological cancer	Dutch	
21	[32]	2014	BPIC 2011	1143	gynaecological cancer	Dutch	
22	[49]	2014	BPIC 2011	1143	gynaecological cancer	Dutch	Other dataset: synthetic log generated using MINERful
23	[24]	2014	BPIC 2011	1143	gynaecological cancer	Dutch	Other datasets: the chest pain patient flow, a commercial insurance claims handling process
24	[48]	2014	two hospitals	258	bronchial lung cancer, colon cancer, rectal cancer, breast cancer, gastric cancer	China	Other dataset: cardiovascular disease - unstable angina
25	[41]	2014	event log of patients	34	rectal cancer	Dutch	in Maastricht University Medical Center (MUMC+)
26	[25]	2015	BPIC 2011	1143	gynaecological cancer	Dutch	
27	[21]	2015	BPIC 2011	1143	gynaecological cancer	Dutch	
28	[30]	2015	BPIC 2011	1143	gynaecological cancer	Dutch	Other datasets: sudden drift case and gradual drift case
29	[22]	2015	BPIC 2011	1143	gynaecological cancer	Dutch	
30	[38]	2015	BPIC 2011	1143	gynaecological cancer	Dutch	
31	[31]	2015	BPIC 2011	1143	gynaecological cancer	Dutch	Other data: BPIC 2012 and BPIC 2014
32	[50]	2015	Data from EMR of Dartmouth-Hitchcock Medical Center	178	breast cancer	New Hampshire USA	Patients with approved chemotherapy agents for invasive breast cancer (Stage I-IV)
33	[51]	2015	Patient data	-	lung cancer	USA	Diagnosis, staging and treatment selection process
34	[60]	2015	therapy processes	40	head and neck cancer	Germany	University Medical Center, Leipzig
35	[63]	2016	dynamic simulation of care pathways	3058	breast cancer	UK	Sepsis (1000) and chemotherapy (3058)
36	[64]	2016	diagnostics from a tertiary referral center	-	gastric cancer	-	Focus on model repair with an example for each of three cases
	[47]	2016	datasets of cervical cancer screening	157	cervical cancer	Italy	Other datasets: the careers of students, data of e-commerce protocol (NetBill)