

Chapter 4

Process Mining in Healthcare



Carlos Fernandez-Llatas, Jorge Munoz-Gama, Niels Martin, Owen Johnson, Marcos Sepulveda, and Emmanuel Helm

4.1 Process Mining

Since medical processes are hard to be designed by consensus of experts, the use of data available for creating medical processes is a recurrent idea in literature [3, 7, 8]. Data-driven paradigms are named to be a feasible solution in this field that can support medical experts in their daily decisions [20]. Behind this paradigm, there are frameworks specifically designed for dealing with process-oriented problems. This is the case of process mining.

Process Mining [32] is a relatively new framework that is thought to provide useful human-understandable information about the processes that are being executed in reality. The process mining paradigm provides tools, algorithms and visualization

C. Fernandez-Llatas (✉)

Process Mining 4 Health Lab – SABIEN – ITACA Institute, Universitat Politècnica de València, Valencia, Spain

CLINTEC – Karolinska Institutet, Sweden

e-mail: cfllatas@itaca.upv.es

J. Munoz-Gama · M. Sepulveda

Pontificia Universitat Católica de Chile, Santiago, Chile

e-mail: jmun@uc.cl; marcos@ing.puc.cl

N. Martin

Research Foundation Flanders (FWO), Hasselt University, Hasselt, Belgium

e-mail: niels.martin@uhasselt.be

O. Johnson

Leeds University, Leeds, UK

e-mail: O.A.Johnson@leeds.ac.uk

E. Helm

University of Applied Sciences Upper Austria, Wels, Austria

e-mail: emmanuel.helm@fh-hagenberg.at

© Springer Nature Switzerland AG 2021

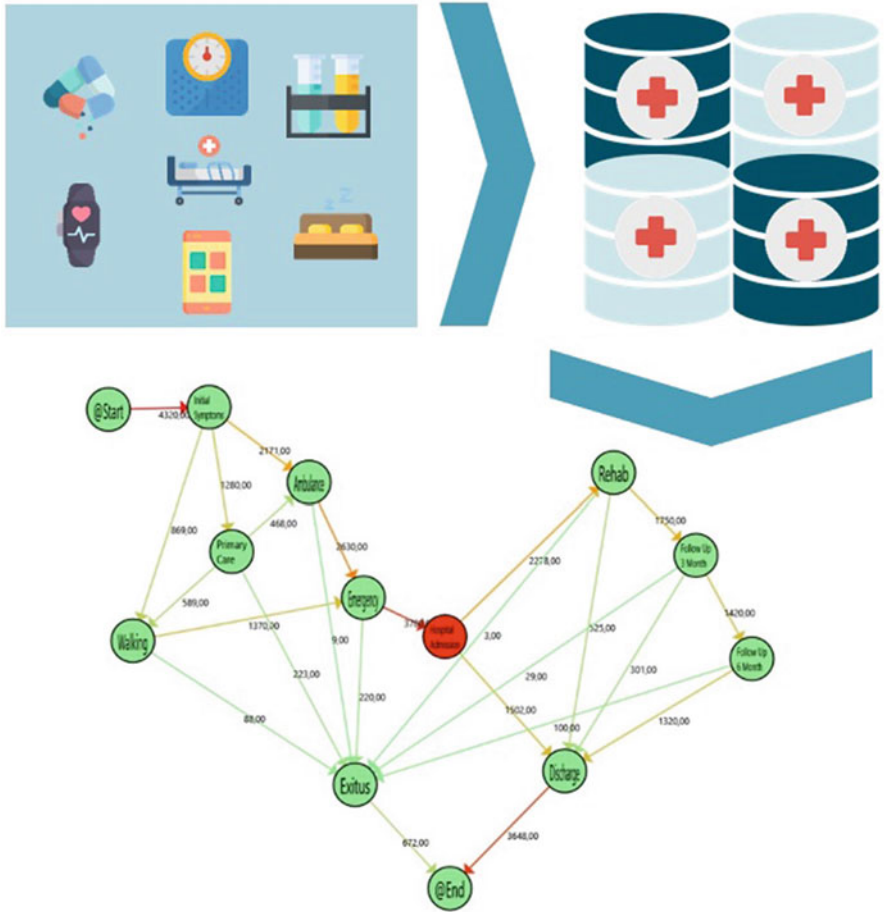
C. Fernandez-Llatas (ed.), *Interactive Process Mining in Healthcare*, Health Informatics, https://doi.org/10.1007/978-3-030-53993-1_4

instruments to allow human experts to obtain information about the characteristics of execution processes, by analyzing the trace of events and activities that occurs in a determinate procedure, from a process-oriented perspective.

Process mining has a close relationship with workflow technologies. Usually, process mining algorithms represent their findings as workflows. Workflows are the most commonly used representation framework for processes. Workflows are not only used in enterprises to automate processes but also clinical guidelines represent some decision algorithms using workflows due to their simplicity and ease of understanding [14]. Besides, the use of formal workflows allows the creation of engines that can automate flows in computer systems. Figure 4.1 shows a possible use of process mining technology. Process mining algorithms use the events recorded in each process and represent it as a workflow. This workflow represents the real flow in an understandable and enriched way, for supporting experts in the actual knowledge of what is occurring in reality. For that, this paradigm can offer a high-level view to professionals, allowing a better understanding of the full process.

Process Mining algorithms are usually divided in three groups [32]

- Process Mining Discovery Algorithms [31]: these are systems that can create graphically described workflows from the events recorded in the process. Different process discovery algorithms are used in several healthcare scenarios [5]. The selection of an adequate discovery algorithm depends on the quantity of data available, and the kind of representation workflow desired. Figure 4.1 shows the inference of a discovery algorithm. This algorithm graphically represents the flow of the patient process from raw data coming from events and activities.
- Process Mining Conformance Algorithms: Process Mining Conformance Algorithms [2] can detect if the flow followed by a patient conforms to a defined process. This can be used to measure the patient's adherence to a specific treatment but allows also the graphical representation of the moments the patient is not fulfilling the treatment flow, supporting the physicians in the improvement process of patients' adherence. These techniques can be also used to compare processes to detect the differences in their executions. Conformance algorithms can compare workflows and show the differences graphically, allowing experts to quickly detect changes in different processes. For example, this technique has been used to detect behavioural changes over time in humans [12].
- Process Mining Enhancement Algorithms: Process Mining Enhancement Algorithms extend the information value of a process model using colour gradients, shapes, or animations to highlight specific information in the workflow, providing an *augmented reality* for a better understanding of the process. For example, Fig. 4.1 shows an example of the enhancement of a workflow. In this case, the workflow is representing the common flow of patients in the surgical area of a hospital [9]. A workflow with colour gradients in nodes is shown, representing the duration of the stay in each stage of the surgical process, and in arrows, representing the change frequency among phases. With this information, experts can have a better idea of the dynamic behaviour of the process and can perform changes to improve the process and evaluate the effectiveness of actions by comparing the current flow with past inferences.



Icons made by Smashicons, Monkik, Itim2101, Pixel Buddha and Freepik from www.flaticon.com

Fig. 4.1 Process mining discovery

The main issue to solve when applying an interactive pattern recognition problem is that the experts should understand the model inferred to provide corrections and infer knowledge from the models [11]. Process mining technology is data-driven but with a focus on understandability. In this line, we can define process mining as a *Syntactic Data Mining technique that supports the domain experts in the proper understanding of complex processes in a Comprehensive, Objective way.* This characteristic makes process mining one of the most suitable technologies for applying interactive models.

4.2 Process Mining in Healthcare

Recent reviews are analyzing the application of process mining techniques in healthcare in detail [5, 27]. Some works analyze the change of hospital processes [27], and management of emergencies [1], support the medical training in surgical procedures [21], the flow of patients in specifically critical departments, like surgery [9], or oncology [26], or even, this framework has been used to analyze the behavioural change in humans to detect early dementia signs [12].

In [5], 447 Process Mining for Healthcare relevant papers were identified. 24 of these papers are indexed on PubMed. PubMed is the most used search engine of the MEDLINE database, containing biomedical research articles offered for the Medicine National Library of the United States. That means that only 5% of the papers that are considered relevant in the Process Mining for Healthcare literature are indexed in the clinical domain. That points to clear difficulties in the application of process mining techniques in real domains. Figure 4.2 shows the trends in publishing in PubMed Library since 2005. This histogram shows that, although there are works about process mining technologies since the start of the century [31], the penetration of these technologies are now starting to be applied in the medical domain.

The difficulties to apply process mining in the healthcare domain are due to distinguishing characteristics of the clinical domain. In this way, it is crucial to take these particularities into account to create a successful process mining system in the health domain:

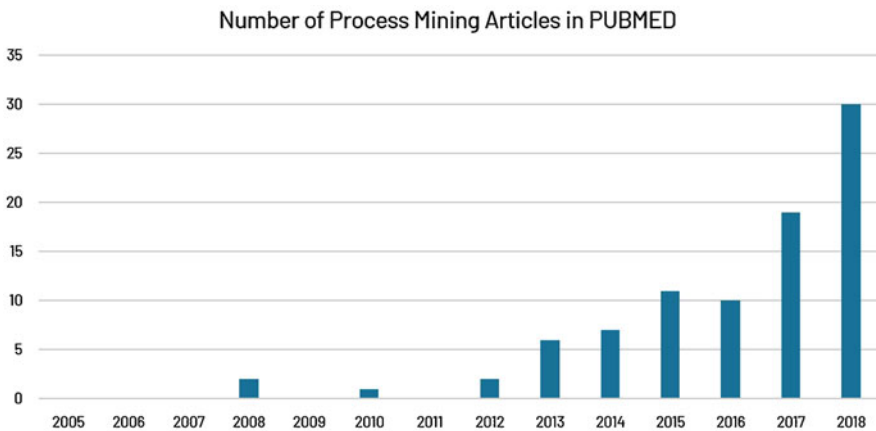


Fig. 4.2 Number of Process Mining articles in PubMed library 2005–2018

4.2.1 *Variability in the Medical Processes*

Medical processes are intended to holistically cover clinical treatments. Medical processes do not only cover diagnosis, treatments, and clinical decisions but also care prevention, patient preferences and medical expertise. That implies that medical processes inherit the complexity of the treatments and diagnosis, as well as the intangible know-how of healthcare professionals [25]. Moreover, patients are usually involved in several health episodes at the same time which causes the co-existence of a different set of pluripathologies and co-morbidities associated to the same patient that can be related, or not, among themselves. This scenario requires close collaboration where many different health professionals interact to define, usually offline, a multi-disciplinary strategy for each patient.

Besides, medical paradigms, like evidence-based medicine [29], value-based healthcare [15], or personalized medicine [16] put the patient in the centre of the medical process. In these paradigms, the attitudes and beliefs of patients are taken into account. This provokes different responses to the treatments, due to the psychology and personal behaviour that take part in the decisions of the patient in terms of the acceptance, or not, of the treatment proposed by the doctor. The adherence of the patient regarding the treatment is one of the most important problems when applying a new treatment to patients [23]. The adherence is key dealing with a disease. The selection of the treatment not only depends on the best option available according to medical evidence, but also on the beliefs, family condition, fears, ambitions, and quality of life of the patient. We should not forget that the real decision-maker in the medical treatment of the patient. In this way, it is crucial to understand their psychological and physical situation before making a clinical decision.

This variability in the medical processes increases the size of the model in terms of arcs and nodes. This provokes one of the most well-known problems in process mining literature, the *Spaghetti Effect*. The application of process discovery algorithms on highly variable systems results in unreadable models. In this way, it is important to select adequate tools for each problem that support the highlighting of the interesting process structures and abstracting or splitting the model in simpler protocols that show the relevant information for the doctors [10]. Process mining researchers should be aware of this problem and provide solutions, tools and frameworks to characterize and extract information about this variability, to extract better and more understandable knowledge of real patients.

4.2.2 *Infrequent Behaviour Could be the Interesting One*

One of the most common solutions that process mining and other data mining practitioners use to infer models is to discard *outliers* in the data. These outliers are considered as noise and are removed from the data. This decreases the variability,

which creates cleaner models that can provide better understandable solutions. However, infrequent behaviour should not always be discarded from the system like, for example, for the detection of adverse effects [30].

Outlier-free logs will produce clean models that produce views representing the most common paths that are followed by the most common patients occurring in most common cases. This implies that the inferred model should be close to the standard clinical processes that are, generally, followed by the patient, and should match with the perception of the medical experts which are providing care. That means that these logs, formed by standard patients, will allow discovering the standard process. However, the standard process is not always the most interesting one. Health professionals usually are familiar with the standard case. The standard patient is covered by the standard treatment. As a consequence, showing the standard model inferred to the doctor does not provide any knowledge to him.

Infrequent cases are cases that do not follow the standard process properly. That means that these cases are the patients that are usually out of the guideline and require special treatment. In those cases, the medical doctor needs help, not as in the standard case. Infrequent behaviour patients processes do not only provide a view about the flow of non-standard cases, but can also provide help to understand the different patient circuits, or even find non-standard similar patients that have followed the same path [22]. This can be a real support to health professionals in daily practice that can't be offered by standard noise reduction techniques.

4.2.3 Medical Processes Should be Personalized

Since the appearance of the evidence-based medicine paradigm, there have been attempts to automate the care provided to patients [29]. The idea of this paradigm is to discover the best medical protocols that can be applied to take care of the disease not only in terms of treatment but also in terms of diagnosis and prevention. However, evidence-based medicine detractors criticize the lack of flexibility in the definition of these protocols [13, 28]. Standard protocols produced by evidence-based medicine has been perceived as incompatible with patient-centred care [28]. Patient-centred care pleads for a more personal application of health where individual's health beliefs, values and behaviours are taken into account when deciding on the best treatments.

All of this criticism on evidence-based medicine originates from the difficulties in the application of clinical pathways due to different health deployment cultures existing in health systems [6, 24]. That means that the application of health care protocols to tackle the same illness can differ significantly from one centre to another, depending not only on the culture of the local population but also on time constraints, the level of staff involvement, the costs associated, among a huge quantity of different factors [6]. Consequently, to replicate the best practices based on medical evidence, it is crucial to iteratively adapt medical processes based on continuous analysis and refinement of the deployed protocols, taking into

account the cultural differences of the target population. Process mining techniques should provide tools and paradigms that allow evolving iteratively the clinical protocols using with the real information collected from target scenarios and led by health professionals. This should be done not only for discovering the standard processes but also supporting health experts in the understanding of the personal characteristics of individual patients.

4.2.4 Medical Processes Are Not Deterministic

Differences in personal preferences of patients, their beliefs and attitudes affect the effectiveness and efficacy of the treatments. That means that in healthcare, cause-effect relationships are often fuzzy. Medical evidence is based on estimations extracted from clinical trials that usually works in most of the cases. However, The same treatment, provided to two different patients with the same illness and, even, the same co-morbidities can result in totally different processes due to many additional factors that cannot be observed or taken into account in the model. This is because there is a gap between each patient and the medical model that represent the clinical knowledge. Formal medical processes are supposed to be automatable and unambiguous, but this is incoherent with the intrinsic nature of patients flow.

This uncertainty is critical in data-driven models. While knowledge systems can offer semantic explanations and rules for describing the ambiguities. Data-driven models can only offer statistical approaches that can inform about the probabilities. The incorporation of process mining techniques, that allow dealing with in-determinism in medical models, like the inference of semantics [4] or the incorporation of information that can point to the ambiguity reasons should be taken into account for better support of the medical decision.

4.2.5 Medical Decisions Are Not Only Based on Medical Evidence, But Also on Medical Expertise

Although one of the most key features of evidence-based medicine is to extract medical evidence, the final decision in medical treatments is always taken by the health professional. Evidence-based medicine looks for the fusion of the best medical evidence with the personal knowledge of medical professionals. This means that medical processes are, in fact, guidelines, that might be followed by the physician. Clinical guidelines aim to offer support to healthcare professionals, based on insights from accumulated clinical evidence.

Data driven systems should focus on providing the most relevant information in the most understandable way to support the practitioner in their daily practice. The decisions taken by health professionals are determined by medical knowledge,

represented by guidelines proposed in the medical models, and their personal feelings obtained from the communication with the patient, which can be verbal or not. The validation of these decisions should aim to provide a holistic value-based healthcare analysis in order to provide medical indicators (like P-Value [18]) that allow professionals to trust the tools provided. These decisions, which can be different from the ones recommended by the models, are registered in the logs and can, consequently, be used in next iterations to improve the models through interactive models [11].

4.2.6 *Understandability Is Key*

Given the difficulties of the manual definition of medical protocols, pathways and guidelines, data-driven technologies are being called to support medical professionals in their formalization. Traditional machine learning techniques are thought to provide the best accuracy in the models, but the interaction with professionals is not available due to the lack of readability of these techniques. Consequently, the resulting models are *black boxes* for health professionals. Machine learning inferred models are based on statistical mathematical models. In this line, the more patients we have in the dataset, the more precise models we can infer [19]. On the contrary, the fewer patients we have the more probable is that the system fails in their prediction. For that, machine learning models are accurate in the prediction of standard cases, but, have a higher probability to fail in infrequent cases. The standard case is usually covered by standard treatment and, in those cases, the expert does not need support. The expert needs support for infrequent cases, where the machine learning models have more problems. This creates suspicion amongst the experts whether to trust a system that has a higher probability of failure for the cases when they are more needed. This is pushing to a new way of explainable machine learning for creating understandable models for health professionals [17].

In the healthcare domain, process mining algorithms should be thought to maximize the understandability of the models inferred. While classical data mining solutions are intended to provide accurate models, process mining enables health professionals a better understanding of the processes and the correction of the models based on their knowledge. Understandable data-driven systems, like process mining, allow the expert to understand the reasons behind the system's recommendation, providing clues for better decisions in daily practice. Besides, it is key to select the correct graphical models that allow the understanding of the processes. Highly expressive models, like Petri nets, can be too complex for the understanding of non-process experts like clinicians, that are used to define their models with more graphically oriented systems as was explained in the previous chapter.

4.2.7 Must Involve Real World Data

Due to the necessity to preserve the privacy of individuals, current laws impose a high barrier for the creation of adequate data-driven models. One potential solution is the use of simulated data for the creation of algorithms to create healthcare solutions. However, the complexity of simulation models is bounded by their model representation capability, which is always much lower than complexity in the real world. Great algorithms and tools that have demonstrated impressive efficacy in other application contexts, can be pointless in healthcare due to unexpected aspects due to unknown variability in the health domain that cannot be simulated. Simulated data cannot offer medical evidence, so, we cannot assume that techniques are adequate for health if we have not tested them with real data.

4.2.8 Solving the Real Problem

The medical expert is the only one able to notice if discovered evidence is relevant or not. Data scientists can find impressive results by creating algorithms to highlight specific aspects of a process. However, if these results do not tackle a real medical problem, or the discovered evidence is well known in the medical community, it makes no sense to use these algorithms. The involvement of medical professionals in the definition of problems and the interpretation of results is key in the creation of useful tools and the discovery of new medical evidence.

4.2.9 Different Solutions for Different Medical Disciplines

Medicine is a huge field formed by several disciplines and specialities having very different variables to measure. Health managers should have different views of the process than clinicians, but even in a single clinical domain, different specialities should have different views. For example, the key biomedical variables for an endocrinologist can be different than the information relevant for a cardiologist. The information available in the medical domain is so huge that it is mandatory to provide the adequate tools and views for each problem to solve, taking into account the real needs of the health professional at any moment. The information overload provokes a paradox effect: the more information is accessible, the more difficult it is to find the relevant one.

For that, the application of process mining technologies in the healthcare domain should be completely adapted to the medical field, creating customized methods and tools for supporting physicians in their daily practice, avoiding one-fits-all solutions and creating adaptive and customizable frameworks for facilitating their real use and highlighting the relevant information in each case.

4.2.10 Medical Processes Evolve in Time

Medical processes are being improved every day due to the continuous appearance of new clinical evidence in the literature. Also, as patients are humans, their personality, beliefs and general behaviour evolve in time. That means that the treatment response of a patient changes depending on several factors that can or cannot be observable. A change in a medical process might suppose an indeterminate change in the behaviour of the process in the next iteration. This is because the effects of change in medical protocols are dependent on an unknown number of variables that makes the result indeterministic.

In this way, process mining technologies should be not only focused on discovering better processes for taking care of the patients but also to be constantly aware of their continuous evolution. This requires tools to trace, measure and analyze how patients adapt their life to the proposed treatments in each stage of their disease. Within a process mining approach, the process can be continuously tracked iteratively. In each one of the iterations of creating optimized and adapted medical protocols, the experts can understand and correct the processes, allowing to be resilient to the concept drift problem produced by the evolution in time of medical processes via the utilization of interactive methods. This is because the interactive paradigm ensures the convergence in the limit of the learning process by involving the expert in the loop [11].

4.3 Conclusion

In the domain of data-driven technologies, process mining has acquired a certain prominence in case of process oriented problems. Its capability to discover, analyze and enhance graphical processes in an easy to understand way supposes a new way to provide information to professionals by involving them in the process of generating knowledge. This characteristic is especially valuable in healthcare, were the expert usually has no engineering knowledge and no data science skills. Unlike other data science paradigms, process mining can provide information to the expert about what is actually occurring with their patients, allowing them in a better understanding of the effectiveness of the treatments selected. This facilitates close collaboration between the computer and the experts that might enable the imbrication between the clinical evidence and the professional knowledge that is required in the evidence-based medicine paradigm.

However, to ensure the applicability of process mining technologies to the clinical domain, it is necessary to take into account their special characteristics. Selected process mining algorithms, methods and tools should be specifically designed to deal with this highly demanding field. In this chapter, we have analyzed and stated the most important of these specificities. The selection of the best

process mining technologies for each specific case is crucial for creating successful deployments of intelligent systems in health centres.

References

1. Andrews R, Wynn MT, Vallmuur K, Ter Hofstede AHM, Bosley E, Elcock M, Rashford S. Leveraging data quality to better prepare for process mining: an approach illustrated through analysing road trauma pre-hospital retrieval and transport processes in queensland. *Int J Environ Res Public Health*. 2019;16(7):1138.
2. Carmona J, van Dongen B, Solti A, Weidlich M. *Conformance checking*. Springer; 2018.
3. Chowriappa P, Dua S, Todorov Y. Introduction to machine learning in healthcare informatics. In: *Machine learning in healthcare informatics*. Springer; 2014. p. 1–23.
4. De Medeiros AKA, van der Aalst WMP. Process mining towards semantics. In: *Advances in web semantics I*. Springer; 2008. p. 35–80.
5. Erdogan TG, Tarhan A. Systematic mapping of process mining studies in healthcare. *IEEE Access*. 2018;6:24543–67.
6. Evans-Lacko S, Jarrett M, McCrone P, Thornicroft G. Facilitators and barriers to implementing clinical care pathways. *BMC Health Serv Res*. 2010;10(1):182.
7. Fatima M, Pasha M, et al. Survey of machine learning algorithms for disease diagnostic. *J Intell Learn Syst Appl*. 2017;9(01):1.
8. Fernández-Llatas C, García-Gómez JM. *Data mining in clinical medicine*. Springer; 2015.
9. Fernandez-Llatas C, Lizondo A, Monton E, Benedi J-M, Traver V. Process mining methodology for health process tracking using real-time indoor location systems. *Sensors (Basel, Switzerland)*. 2015;15(12):29821–40.
10. Fernandez-Llatas C, Martinez-Millana A, Martinez-Romero A, Benedi JM, Traver V. Diabetes care related process modelling using process mining techniques. Lessons learned in the application of interactive pattern recognition: coping with the Spaghetti Effect. In: *2015 37th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2015. p. 2127–30.
11. Fernandez-Llatas C, Meneu T, Traver V, Benedi J-M. Applying evidence-based medicine in telehealth: an interactive pattern recognition approximation. *Int J Environ Res Public Health*. 2013;10(11):5671–82.
12. Fernández-Llatas C, Benedi J-M, García-Gómez JM, Traver V. Process mining for individualized behavior modeling using wireless tracking in nursing homes. *Sensors*. 2013;13(11):15434–51.
13. Goldberger JJ, Buxton AE. Personalized medicine vs guideline-based medicine. *Jama*. 2013;309(24):2559–60.
14. Gooch P, Roudsari A. Computerization of workflows, guidelines, and care pathways: a review of implementation challenges for process-oriented health information systems. *J Am Med Inform Assoc*. 2011;18(6):738–48.
15. Gray M. Value based healthcare. *BMJ*. 2017;356:j437.
16. Hamburg MA, Collins FS. The path to personalized medicine. *N Engl J Med*. 2010;363(4):301–4.
17. Holzinger A. From machine learning to explainable AI. In: *2018 world symposium on digital intelligence for systems and machines (DISA)*. IEEE; 2018. p. 55–66.
18. Ibanez-Sanchez G, Fernandez-Llatas C, Celda A, Mandingorra J, Aparici-Tortajada L, Martinez-Millana A, Munoz-Gama J, Sepúlveda M, Rojas E, Gálvez V, Capurro D, Traver V. Toward value-based healthcare through interactive process mining in emergency rooms: the stroke case. *Int J Environ Res Public Health*. 2019; 16(10). <https://www.mdpi.com/1660-4601/16/10/1783>.

19. Jain AK, Chandrasekaran B. 39 dimensionality and sample size considerations in pattern recognition practice. *Handb Stat.* 1982;2:835–55.
20. Kononenko I. Machine learning for medical diagnosis: history, state of the art and perspective. *Artif Intell Med.* 2001;23(1):89–109.
21. Lira R, Salas-Morales J, Leiva L, Fuentes R, Delfino A, Nazal CH, Sepúlveda M, Arias M, Herskovic V, Munoz-Gama J et al. Process-oriented feedback through process mining for surgical procedures in medical training: the ultrasound-guided central venous catheter placement case. *Int J Environ Res Public Health.* 2019;16(11):1877.
22. Mansour A, Ying H, Dews P, Ji Y, Yen J, Miller RE, Massanari RM. Finding similar patients in a multi-agent environment. In: 2011 annual meeting of the North American fuzzy information processing society. IEEE; 2011. p. 1–6.
23. Osterberg L, Blaschke T. Adherence to medication. *N Engl J Med.* 2005;353(5):487–97.
24. Pace KB, Sakulkoo S, Hoffart N, Cobb AK. Barriers to successful implementation of a clinical pathway for CHF. *J Healthcare Qual Off Publ Natl Assoc Healthcare Qual* 2002;24(5):32–8.
25. Rebuge Á, Ferreira DR. Business process analysis in healthcare environments: a methodology based on process mining. *Inf Syst.* 2012;37(2):99–116.
26. Rinner C, Helm E, Dunkl R, Kittler H, Rinderle-Ma S. Process mining and conformance checking of long running processes in the context of melanoma surveillance. *Int J Environ Res Public Health.* 2018;15(12):2809.
27. Rojas E, Munoz-Gama J, Sepúlveda M, Capurro D. Process mining in healthcare: a literature review. *J Biomed Inform.* 2016;61:224–36.
28. Romana H-W. Is evidence-based medicine patient-centered and is patient-centered care evidence-based? *Health Serv Res.* 2006;41(1):1.
29. Straus SE, Richardson WS, Glasziou P, Haynes RB. Evidence-based medicine: how to practice and teach EBM. 3 ed. Churchill Livingstone; 2005.
30. Stricker BHCh, Psaty BM. Detection, verification, and quantification of adverse drug reactions. *BMJ.* 2004;329(7456):44–7.
31. Van der Aalst W, Weijters T, Maruster L. Workflow mining: discovering process models from event logs. *IEEE Trans Knowl Data Eng.* 2004;16(9):1128–42.
32. van der Aalst WMP. Process mining: data science in action. Springer; 2016.