

Business Process Mining in Healthcare with State-of-the-Art Open Source Tools

Murat SARIYAR ^{a,1}

^a*Bern University of Applied Sciences, Switzerland*

ORCID ID: Murat Sariyar <https://orcid.org/0000-0003-3432-2860>

Abstract. This study investigates the application of business process mining to optimize patient flow, minimize wait times, and enhance resource utilization in an ophthalmology clinic. Using open-source R packages, particularly bupaR, we analyzed event logs derived from patient data across multiple IT systems, mapping the patient journey from scheduling to discharge. The findings demonstrate that process mining effectively supports the understanding and improvement of organizational workflows. Additionally, the study highlights that commercial tools are not essential for such analyses; individuals with moderate IT expertise can easily adopt open-source alternatives. Crucially, success depends not on mastering the tools' intricate mechanisms but on the ability to accurately interpret results and understand key methodological parameters.

Keywords. Process Mining, Patient Flow, bupaR, R Programming.

1. Introduction

Understanding and analyzing processes are central tasks in medical informatics [1]. While substantial progress has been made in many areas of digitalization, process analysis in healthcare often still relies heavily on manual methods [2]. However, this approach is prone to errors, time-intensive, and limited in its ability to accurately represent complex, dynamic workflows typical of clinical practice. In this context, business process mining is becoming increasingly relevant [3]. This technique enables the automated capture, visualization, and optimization of processes based on digital traces left by information systems. Especially in healthcare, where large volumes of data are generated across various IT systems, process mining has the potential to identify inefficient workflows, enhance patient safety, and improve resource utilization [4].

The utilization of state-of-the-art open-source tools in this domain offers a cost-effective and flexible solution for conducting comprehensive process analyses. These tools not only provide transparent and reproducible results but also integrate seamlessly into existing IT infrastructures [5]. Thus, the application of modern process mining approaches represents a critical step towards data-driven process optimization in healthcare [6]. Conversely, many commercial products, such as *SAP Signavio Process Intelligence* or *Fluxicon Disco*, often rely on open-source frameworks like *ProM* as a foundational basis, primarily focusing on visualization as their added value [7–9].

¹ Corresponding Author, Murat Sariyar, Bern University of Applied Sciences, Quellgasse 21, CH2502 Biel/Bienne, Switzerland; E-mail: murat.sariyar@bfh.ch.

This study presents the application of process mining in an ophthalmology clinic using R packages. The application of process mining using open-source tools is novel in this context, highlighting the obsolescence of commercial solutions. We begin by introducing the use case, followed by a brief discussion of the specific R packages employed and the analytical capabilities they offer. Key results and interpretations are then provided, concluding with practical recommendations for practical implementation.

2. Methods

The primary objective of our use case is to optimize patient flow within the ophthalmology clinic, aiming to minimize wait times, enhance resource utilization, and improve the quality of patient care. Through the analysis of patient data and treatment workflows, we can identify bottlenecks and streamline processes to achieve greater operational efficiency. The clinic utilizes various systems to collect patient information, including scheduling and billing systems. Key data points that need to be collected for process mining include the following items:

- Patient arrival times
- Waiting periods before treatment
- Duration of examinations and treatments
- Patient discharge times

The data are standardized into a unified XES (eXtensible Event Stream) event log format, necessitating a sophisticated ETL (extract, transform, load) process due to the need for extracting and transforming data from multiple systems [10]. Data were extracted from *Ultragenda* (scheduling) and *Opale* (resource management), cleaned for consistency, and transformed into event logs by standardizing timestamps, mapping key events (e.g., scheduling, diagnostics, discharge), and integrating relevant attributes. Using process mining tools, a visual model of the current treatment process is generated, mapping each step in the patient flow from appointment scheduling through to discharge. This model identifies bottlenecks, such as prolonged wait times, and detects deviations from established process standards. These analytical insights inform the development of targeted process optimization measures. Our study focuses on identifying inefficiencies, leaving the exploration of specific optimization measures for future research.

We utilize *bupaR* as our process mining tool – an open-source R package specifically developed for business process analysis [11]. The package provides an extensive suite of functions, including the generation of event log objects that enable filtering of event data, calculation of descriptive statistics, and visualization of process maps. Core and related packages supporting *bupaR* include:

- *edeaR*: designed for exploratory data analysis and data preprocessing, facilitating the preparation and examination of datasets for subsequent analysis.
- *processmapR*: used for visualizing processes to aid in process analysis and serves as a core component of the *bupaR* framework.
- *eventdataR*: provides a repository of sample data, enabling users to learn and apply the functionalities of the package in a practical context.

With *bupaR*, users can not only analyze processing times and activity frequencies but also capture a range of relevant metrics that provide valuable insights into the efficiency and effectiveness of their business processes [12].

3. Results

The data were extracted from various systems using SQL, specifically focusing on patient admissions, triage, evaluations, consultations, and discharges. The extraction process involved retrieving timestamped events essential for process mining. Subsequently, the data were pre-processed, and the event log was generated. This log serves as the foundational dataset for all subsequent analyses. Notably, the creation of the process map is central to this analysis, as it provides a comprehensive overview of the process steps. The diagram in Figure 1 was generated using the following sequence of commands, which illustrates the ease of use of the package

```
event_log %>%
  filter_infrequent_flows(min_n = 1000) %>%
  process_map()
```

The code generates a process map from the event log, filtering out most of the infrequent flows (edges) to ensure that only common and relevant sequences are displayed in the process map. This approach enhances clarity, which is crucial given that the dataset contains a total of 1,364 traces, resulting in a considerable number of diverse and often unanticipated pathways through the system.

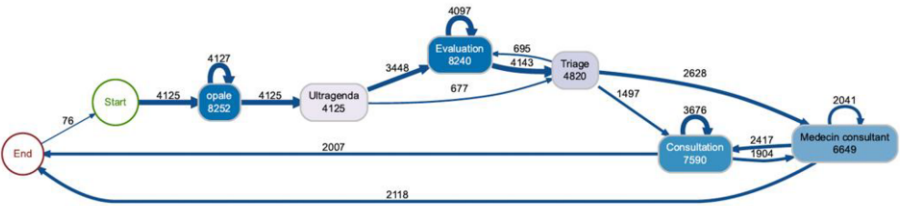


Figure 1: Process map for the procedures at the ophthalmology clinic, as provided by bupaR.

The process map provides a visual representation of the most frequent pathways for patients, starting from the initial contact with the healthcare system to the final consultation with a physician. For the most prevalent cases, the trajectory involves progression through the Opale platform to the appointment scheduling tool, followed by evaluation and subsequent consultation or treatment. The processes of evaluation and triage are tightly interwoven, forming a single critical phase that serves as a decisive juncture. Here, the patient's condition is assessed, and care priorities are established to ensure timely and appropriate attention to urgent cases. This integrated phase is essential for directing immediate next steps in the care pathway. Similarly, the roles of the consultant physician and the consultation phase are closely aligned, marking the point where a comprehensive examination occurs. At this stage, the physician integrates clinical findings with the patient's medical history and symptoms to establish a diagnosis and develop a treatment plan. This step is pivotal, as it often determines subsequent care actions, such as further testing, hospital admission, or discharge with follow-up instructions. Notably, the map highlights iterative loops, indicating processes that may need to be repeated before progression. These loops suggest areas where inefficiencies in the system could lead to delays or redundancies in patient care.

To improve the current system and minimize unnecessary procedural redundancies, several strategic improvements can be implemented. First, increasing data capture consistency can be achieved by establishing a comprehensive protocol for data entry. Second, the workflow could be optimized by consolidating redundant steps, particularly

by merging 'evaluation' and 'triage' processes where feasible. Third, staff education can effectively reduce confusion about the interactions among IT systems. By clarifying the rationale behind each procedural step, it promotes a better understanding of individual roles and enhances satisfaction within a streamlined operational framework.

Figure 2 illustrates the aggregated average waiting times by hour for central workflows. The y-axis indicates the total waiting time in minutes, while the x-axis represents the hours of the day. The workflows are identified by their abbreviations: O-E-C (Opale-Evaluation-Consultation), O-E-M (Opale-Evaluation-Med. Consultation), O-T-C (Opale-Triage-Consultation), and O-T-M (Opale-Triage-Med. Consultation). Analysis of the trends reveals that waiting times for all workflows consistently increase throughout the day, peaking notably in the late evening. This pattern suggests insufficient management of patient influx during morning hours, resulting in a cumulative buildup of delays. These findings highlight opportunities for optimizing shift scheduling.

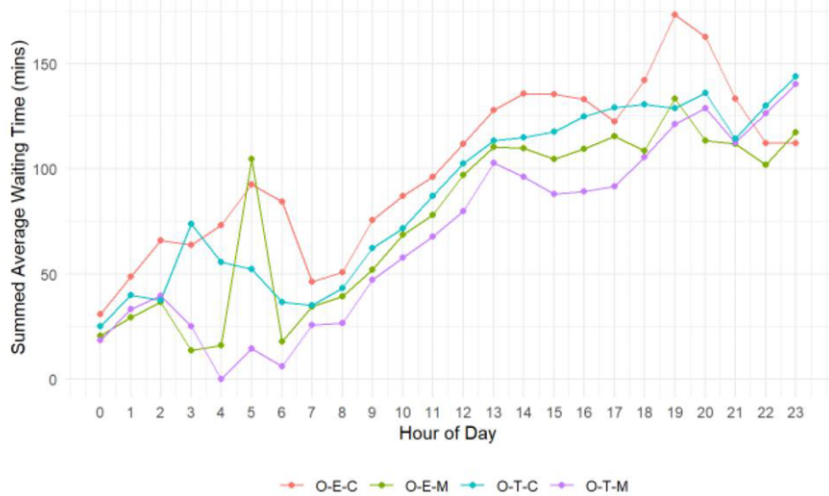


Figure 2: Aggregated average waiting times by hour for central workflows indicated by colors.

4. Discussion

Our results identify a primary pathway with feedback loops, suggesting inefficiencies. Constructing the system requires a solid understanding of R programming, process mining principles, and workflow dynamics. Additionally, the analysis offers new insights into patient processing, highlighting that:

- The process can be optimized by combining steps or eliminating redundancies, such as "Assessment" and "Triage."
- Staff should receive training to enhance their understanding of collaboration between various IT systems and the roles involved in the process.
- Shift scheduling should be reevaluated to ensure that workflow transitions smoothly from the morning to the later parts of the day.

Overall, the project clearly demonstrates that process mining is an effective method for understanding and improving organizational processes, offering valuable insights into the flow of activities and the identification of inefficiencies. It highlights that commercial process mining products are not always necessary to achieve meaningful

outcomes, as open-source or custom-built solutions can also be effective, depending on the specific needs of the organization. Additionally, the project suggests that the onboarding process for using process mining tools is manageable for individuals with a basic level of IT knowledge, especially when the tools are designed to be user-friendly, which also applies to command-line tools like bupar.

However, more crucial than the tool itself is the ability to accurately interpret its results. Understanding the underlying algorithms, such as HeuristicMiner, and knowing how to adjust parameters to match the specific context of the organization is paramount. This does not require in-depth technical expertise in the method's implementation, but rather a solid understanding of the principles behind the analysis, as well as the ability to translate findings into actionable improvements. Our approach is highly adaptable to diverse clinical settings, accommodating variations in patients and workflow dynamics.

References

- [1] De Ramón Fernández A, Ruiz Fernández D, Sabuco García Y. Business Process Management for optimizing clinical processes: A systematic literature review. *Health Informatics J* 2020; 26: 1305–1320.
- [2] Lenz R, Reichert M. IT support for healthcare processes – premises, challenges, perspectives. *Data Knowl Eng* 2007; 61: 39–58.
- [3] Munoz-Gama J, Martin N, Fernandez-Llatas C, et al. Process mining for healthcare: Characteristics and challenges. *J Biomed Inform* 2022; 127: 103994.
- [4] Guzzo A, Rullo A, Vocaturo E. Process mining applications in the healthcare domain: A comprehensive review. *WIREs Data Min Knowl Discov* 2022; 12: e1442.
- [5] Tornero-Costa R, Fernandez-Llatas C, Martin N, et al. From Script to Application. A bupaR Integration into PMAApp for Interactive Process Mining Research. In: Juarez JM, Fernandez-Llatas C, Bielza C, et al. (eds) *Explainable Artificial Intelligence and Process Mining Applications for Healthcare*. Cham: Springer Nature CH 2024, 107–117.
- [6] Potts C, Bond RR, Jordan J-A, et al. Process mining to discover patterns in patient outcomes in a Psychological Therapies Service. *Health Care Management Science* 2023; 26: 461–476.
- [7] Scheruhn H-J, Reiter C, Bayramli E. From Process Mining to Enterprise Mining. In: Marx Gómez J, Elikana Sam A, Godfrey Nyambo D (eds) *Artificial Intelligence Tools and Applications in Embedded and Mobile Systems*. Cham: Springer Nature Switzerland, 2024, pp. 195–205.
- [8] Rabbi F, Lamo Y, MacCaull W. A Model Based Slicing Technique for Process Mining Healthcare Information. In: Babur Ö, Denil J, Vogel-Heuser B (eds) *Systems Modelling and Management*. Cham: Springer International Publishing, 2020, pp. 73–81.
- [9] Rashed A-HM, El-Attar NE, Abdelminaam DS, et al. Analysis the patients' careflows using process mining. *PLOS ONE* 2023; 18: e0281836.
- [10] Kampik T, Weske M. Event Log Generation: An Industry Perspective. In: Augusto A, Gill A, Bork D, et al. (eds) *Enterprise, Business-Process and Information Systems Modeling*. Cham: Springer International Publishing, 2022, pp. 123–136.
- [11] Van Hulzen G, Janssenswillen G, Martin N, et al. Process Analysis with bupaR 0.5.0: What's New?, <https://documentserver.uhasselt.be/handle/1942/39066> (2022, accessed 25 October 2024).
- [12] Chen L, Klasky HB. Six Machine-Learning Methods for Predicting Hospital-Stay Duration for Patients with Sepsis. In: *Southeast Con 2022*, pp. 302–309.