

Article

# A Goal-Driven Evaluation Method Based On Process Mining for Healthcare Processes

Tugba Gurgen Erdogan \* and Ayca Tarhan

Computer Engineering Department, Hacettepe University, 06532 Ankara, Turkey; atarhan@hacettepe.edu.tr

\* Correspondence: tugba@cs.hacettepe.edu.tr; Tel.: +90-312-297-7500

Received: 7 April 2018; Accepted: 27 May 2018; Published: 30 May 2018



**Abstract:** As a business processes management technique, process mining (PM) has been applied in many domains in the last decade. In healthcare, where most processes are complex, variable, dynamic, and multi-disciplinary in nature, application of this technique is growing, yet challenging. Therefore, this study aims to introduce a goal-driven process evaluation method based on PM for healthcare processes. The proposed method comprises the following steps: defining goals and questions, data extraction, data preprocessing, log and pattern inspection, PM analysis and generating answers to questions, evaluating results, and initiating proposals for process improvements. The proposed method was applied in a case study on the surgery process of a university hospital in Turkey, which revealed for quantitative insights into the process. Bottlenecks and deviations that were crucial for determining measures (e.g., data and performance information) were identified to improve the efficiency of the surgery process. Our initial experience using the proposed method shows that it has potential for initiating process improvements by guiding the use of PM techniques in the healthcare domain.

**Keywords:** healthcare process; clinical pathway; process mining; process management; evidence-based medicine

## 1. Introduction

In healthcare process management, healthcare professionals aim to formalize, improve and optimize processes by creating better clinical pathways (CPs). A CP is one of the main tools used to manage quality of healthcare and guide evidence-based medicine [1]. Medical professionals should develop pathways in an objective manner and when putting them into practice, they should deal with variances and individual care, adaptation to target population, concept drifts, and new healthcare paradigms. Using CPs in daily practices, however, should be addressed only with the right tools and techniques [2]. Considering that almost every step of clinical processes is undertaken digitally and recorded (as event logs) in information management systems, healthcare data can provide insights into the weaknesses of the system to suggest improvements. Refining the definition of CPs by exploiting healthcare data may be an appropriate way of managing healthcare processes for medical professionals.

Data-driven CPs are generated by applying data mining, machine learning or statistical modeling in order to respond to performance-oriented and compliance-oriented questions of CP analysis at run-time and design time. At this point, process mining (PM) is a powerful, objective technique that has been applied in various domains including healthcare. This technique links model-based process and data-oriented analysis. In the healthcare domain, application of this technique is promising due to the growing number of reported studies [3]. PM can provide several operational insights into healthcare processes: concerning process flows, checking conformance of data with recommendations, identifying performance and waiting times in processes, and managing resource allocation. However, despite the

body of evidence on the applicability and usefulness of PM techniques in the healthcare domain, there are still significant methodological and data quality challenges that need to be addressed.

Based on the need stated above and a recent literature review conducted by the same researchers [4], this article proposes a novel, PM-based process evaluation method for improving healthcare processes. This goal-oriented method aims to (1) provide a broad overview of processes in a short time, (2) create improvement opportunities for processes by incorporating domain knowledge into process enactments, (3) analyze unstructured, multidisciplinary processes like healthcare processes, (4) implement PM activities and techniques in PM methodology, and (5) provide a practical guidance for traceability, usability and functionality of PM implementations using a goal-driven approach. This method starts with the definition of the scope and creation of a table called GQFI that addresses initial project goals (G), questions (Q), features (F) and performance indicators (I) of a PM project. In the following step, event data extraction is performed based on goals and questions. Then, data is preprocessed (by modeling, filtering or abstracting) to provide a framework for interpreting it with respect to the stated goals. Following the GQFI table, the PM analysis stage is executed in which performance indicators selected for each question are measured. The results (indicators and answers to questions) are used as inputs to the following stage of evaluating process performance. Finally, process improvement opportunities are discussed and identified.

The research described in this paper was undertaken by following a design science approach [5] in which artifacts were built and evaluated to meet the identified business needs [6]. In this paper, an artifact refers to a goal-driven evaluation method, and the research is divided into two stages: a design and build stage, and an evaluation stage for the constructed validation of the artifact through a case study. The feasibility of the proposed method was demonstrated with an embedded and single case study [7] involving the surgery process, and the results showed that the proposed method has potential in initiating proposals.

The remaining of this article is organized as follows: Section 2 presents background information on evidence-based medicine and the role of PM in this area. Section 3 provides an overview of the generic or healthcare specific PM methodologies. Section 4 gives detailed explanation about the PM-based process evaluation method. Section 5 contains a practical example from a case study on surgery process and its results. Section 6 discusses the results of the case study to identify opportunities for future research. The last section presents the overall conclusions and recommendations for future work.

## 2. Background

PM is an objective and useful process management technique [8] and has been applied in various domains including healthcare. In this section, we provide a basic overview of healthcare processes, evidence-based medicine and the role of PM in healthcare.

### 2.1. Healthcare Process

Healthcare processes include different activities by the cooperation or consultation of different organizational units and medical disciplines and based on knowledge and information. Healthcare processes are highly dynamic, complex, increasingly multi-disciplinary and ad hoc [9]. Activities performed are either administrative, such as admission, registration, transferring to service, and discharge, or are medical, such as diagnosis, operation, tests and scans, and intervention. Healthcare processes are classified as organizational processes and medical treatment processes [10]. Organizational processes have a repetitive nature and do not provide any support for medical decision making. The tasks of physicians, nurses, and technicians in process are defined as organizational processes and support the medical treatment processes which are influenced by medical knowledge and patient related information. Medical treatment processes start with patient history and continue with diagnostic procedures required medical decisions. Medical treatment processes may become highly unpredictable and complex based on individual cases.

## 2.2. Evidence-Based Medicine

The evidence-based medicine paradigm introduced in the last decade represents the integration of individual clinical expertise, the patient's values and expectations, and the best available evidence in the process of decision-making for the patient's healthcare [11,12]. In this process, it is recommended to use CPs and clinical practice guidelines [13], which help improve clinical expertise with better evidence [14]. Furthermore, clinical guidelines, protocols, and pathways support quality improvement and cost control activities of care and formalize the best medical practices. Clinical guidelines provide systematically developed statements and assist clinicians in making decisions [15]. Clinical pathways are defined as a set of therapy and treatment activities that represent the steps required to achieve a specific treatment objective for a particular disease [16]. Pathways are a distinctive type of clinical guidelines, which specify each step in the care process, rather than stating broad principles that practitioners should follow [17]. Nevertheless, it was observed that there is a gap between these recommendations and actual practices since healthcare processes have large healthcare facilities, multidisciplinary and highly complicated structures, and flexible patient flows [18,19].

Using clinical recommendations in daily practices is a crucial challenge in patient-oriented, evidence-based medicine paradigm for healthcare process management. The definition of a healthcare process should be based on actual implementations for detecting the best practices and removing bad practices (by observing outliers) rather than assumed workflows.

## 2.3. Process Mining

PM is relatively young discipline used to exploit event data on information systems for (i) process discovery, (ii) conformance checking of process implementations according to the discovered/ modeled process, and (iii) enhancement of process by detecting the differences in process implementations [8]. Event logs store detailed information on events concerning the source; i.e., the person or tool that started and performed the activity, the starting and finishing time of the activity, and the data element; e.g., type, size, and comments [8].

Process discovery is the most important activity of PM since it provides a basis for further analyses involving the application of the remaining two types of PM; conformance checking and enhancement. In process discovery, event logs are used as an input, and a process model is set up without prior information [8]. When an actual process is created from event logs, many organizations can face challenges due to the differences in theory and implementation.

In conformance checking, the process model and its flow discovered from the event logs are analyzed, and it is checked whether the process has been carried out as identified in the model [8]. Conformance checking measures the differences between the performed process and the process model specifications. The main aim of this activity is to identify the areas that need improvement using the information gained from the actual process [20].

Process enhancement is the improvement of the process model based on event data. This can be undertaken by adding further data perspectives to the process model using event data, an activity also known as extension. Another type of enhancement is repair, in which the quality of the process model is improved using event data to define a new repaired model [19].

Through process exploration techniques, a process model is discovered from event logs and is evaluated. This process may be repeated several times until a satisfying model is achieved. The activity model can also be refined manually or automatically to obtain better representations with process enhancement techniques after the automatic process exploration activity. It also allows checking the conformity between actual clinical processes and recommendations. In addition, variations of recommendations can be improved using interactive visualization techniques and user-defined levels of abstraction. Selecting suitable representation notation and providing a visualization environment are pre-requirements for defining recommendations.

Various software products have PM capabilities [21]; e.g., ARIS Process Performance Manager (Software AG) [22], ProM (TU/e) [23], Disco (Fluxicon) [24], PALIA-ER [25], CELONIS [26], and

pMineR [27]. Some of these products are still at the maturing stage since PM is a relatively young paradigm open to research and practice. For example, ProM [23] is a basic PM tool, which provides a standard environment incorporating a generic open-source framework for implementing PM techniques. Disco [24] is a popular, powerful, fast, and easy-to-use PM toolkit. The revolutionary commercial PM technology in Disco helps researchers to create visual maps from process data in minutes. In addition to these generic-purpose PM tools, pMineR introduced as an R library has been designed to perform PM specifically in the medical domain, and now supports two types of PM analysis; process discovery and conformance checking [27]. PALIA-ER is a web-based PM tool designed for question-driven PM analysis of the emergency room domain and includes only process discovery by using the algorithm PALIA [28].

### 3. Related Works

PM has the potential to offer many advantages for healthcare professionals. Particularly in healthcare, where most processes are complex, variable, dynamic, and multi-disciplinary in nature, the application of this technique is growing yet challenging. In this research, first, a descriptive analysis of the PM studies on healthcare [4] was undertaken by applying systematic mapping [29]. A total of 172 relevant papers were analyzed with respect to various aspects including research and contribution type, application context and healthcare specialty, process modeling type and notation, PM activities, PM techniques, and demographic and bibliometric analysis. The results of mapping showed that, despite the challenges related to healthcare data and technique, this field is rapidly growing and open for further research and practice.

The systematic mapping study also showed that PM community in the healthcare domain is lacking a comprehensive methodology that describes how to accomplish PM activities and techniques in practice. In data mining, efforts have been made to establish methodologies to support organizations in their data mining projects [30]. The aim of these methodologies is to guide the planning and execution of such projects in order to save time and cost. The commonly used methodologies include CRISP-DM [31], SEMMA [32], and KDD [33], which are very high-level and provide only little guidance support. Furthermore, none of the available methodologies have been tailored as specific to PM projects [34]. The main difference between data mining and PM methodologies concerns the analysis stage, in which PM considers the whole process in terms of events with the main objectives of process discovery, conformance checking and process enhancement, whereas data mining projects analyze data statistics to determine unsuspected correlations.

During the literature review, four generic PM methodologies were identified: Process Diagnostic Method (PDM) [35], which has also been extended to the healthcare domain [9]; the L\* life-cycle model [8]; PM2: Process Mining Project Methodology [34]; and Business Process Comparison Methodology [36]. PDM was designed to obtain a broad overview of processes; however, its scope is limited and it does not utilize domain knowledge. The L\* life-cycle model covers different distributed aspects of PM techniques, and is designed for structured processes. PM2 guides the execution of PM projects by defining and answering questions iteratively to improve process performance. Business process comparison methodology focuses only on the analysis of multiple processes by comparing processes or process variants.

In the related literature, 12 healthcare specific studies were found to have proposed processes to guide the use of PM techniques for process enhancement. Of these, four [37–40] present their proposals without having them validated, and therefore do not include empirical evidence for applicability or usefulness. Question-driven methodology [41] focuses on the application of PM projects to provide answers to questions about emergency room processes. Some studies [42–44] suggest combining PM and simulation to discover and evaluate as-is and to-be processes, respectively. Yang et al. [45] focus on clustering patient traces and process recommendations and Dagliati et al. [46] emphasize deriving frequent healthcare flows and clinical data enrichment. Rovani et al. [19] address the gap between daily practice and clinical guidelines based on declarative process models, and Fernández-Llatas et al. [47]

provide filtering options and algorithms for process discovery, conformance, and enhancement by implementing PALIA ILS tool suite.

Table 1 presents the relation of methodologies with PM features that need to be addressed for analyzing healthcare processes. Based on the findings and challenges identified in previous work [4], the current research extends PM activity types as PM features by comparing and collecting the steps of generic PM methodologies. Furthermore, indicators that can be measured by using these features were determined and are listed in Table 2. These indicators are collected from the studies which reported quantitative benefits through the use of PM techniques. Both Tables 1 and 2 were created as specific for this study. In Table 1, the contribution of each methodology is evaluated in terms of PM features in comparison to the method proposed in this study. The information in this table reveals the need for a comprehensive methodology for in-depth practical operational guidance to assist PM practitioners in a purposeful way for the management of healthcare-specific processes.

**Table 1.** PM features of existing methodologies and the proposed method.

PM Features and Methodologies	PDM [35]	L* Lifecycle Model [8]	PM2 Project [34]	Process Comparison Methodology [36]	Question Driven Methodology [41]	Declarative PM [19]	The Proposed Method
Defining Goals		✓	✓				✓
Defining Questions		✓	✓		✓		✓
Defining Indicators							✓
Selecting PM Features							✓
Designing de jure model		✓				✓	✓
Data Extraction	✓	✓	✓		✓		✓
Data Preprocessing	✓		✓	✓	✓		✓
Data Modeling							✓
Data Abstraction			✓				✓
Data Filtering			✓		✓		✓
Data Visualization			✓		✓		✓
Clustering				✓			✓
Log and Pattern Inspection	✓		✓		✓		✓
Process Discovery	✓	✓	✓	✓	✓	✓	✓
Process Exploration		✓				✓	✓
Process Variant Analysis				✓	✓		✓
Conformance Checking	✓	✓	✓		✓	✓	✓
Process Enhancement		✓	✓			✓	✓
Decision Point Analysis							✓
Organizational Analysis	✓				✓		✓
Performance Analysis			✓		✓		✓
Performance Target Analysis							✓
Process Comparison				✓			✓
Process Improvement		✓	✓				✓
Operational Support		✓	✓				✓

The proposed method aims to monitor, control and improve process performance and is based on a goal-driven model to incorporate domain knowledge through an iterative analysis and through the practice of distributed PM. The GQFI table provides guidance for the interpretation of performance and compliance, and allows eliciting improvement opportunities based on bottlenecks and inconsistencies as an outcome. This method is suitable for unstructured processes, such as healthcare, and provides a guide for the state-of-the-art PM techniques, extension of data preprocessing steps, data visualization techniques, process exploration, and outlier detection. It helps achieve project goals such as improve a process with respect to particular key performance indicators (KPIs), obtain concrete data, and frame performance-oriented questions, e.g., especially frequently posed questions (FPQs) [3].

**Table 2.** Indicators that can be measured by PM features.

PM Feature	Indicators
Log and Pattern Inspection	Number of patients/events/activities Mean duration of all patients Start and finish time of process Frequencies and relative frequencies of data attributes Frequencies and relative frequencies of resource attributes Number of variants Duration of each patient
Process Discovery	Identities of process starting/ending activities Number/identities of paths in process map Number of patients/events in each path Number of patients/events in each activity Number of rework loops
Process Exploration Process Enhancement Process Comparison	Total number of variants Cyclicity and concurrency in process models Fitness, precision, generality, simplicity in process models
Process Variant Analysis	Percentage of patients in each variant Traces (sequence of events) in each variant Identities of patients in each variant Mean duration of each variant Number of outlier patients/events in each variant
Conformance Checking	Matching rate Cost-based fitness value Fitting traces/paths Non-fitting traces/paths Identities of cases in fitting/non-fitting traces
Decision Point Analysis	Frequencies and relative frequencies of data attributes Number of decision points Decision trees for each decision point
Organizational Analysis	Frequencies and relative frequencies of resource attributes Handover of work and working together and resource efficiency
Performance Analysis	Total/Mean/Min/Max/Standard Deviation(STD) of duration Total/Mean/Min/Max/STD of duration of each path
Performance Target Analysis	Total/Mean/Min/Max/STD of duration of each activity Total/Mean/Min/Max/STD of waiting time between activities
Outlier Detection	Number and percentage of outlier process variants Number and percentage of outlier patients/events in each variant

#### 4. Method

This section proposes and explains the goal-driven process evaluation method based on PM for improving healthcare processes. The method is based on a refined PM framework, guidelines and methodologies for PM projects [8,34], and adopted goal- and question-driven PM projects [8]. It details the main stages of a healthcare process evaluation project to improve usability and understandability of PM in unstructured processes for non-experts. The proposed PM methodology seeks ways of presenting a process in a simple manner by dividing data into smaller datasets, and focusing on their specific characteristics. An overview of the goal-driven healthcare process evaluation method is shown in Figure 1. This method can deal with a large amount of healthcare data by distributing PM activities and obtaining valuable information in the following steps: (1) The scope and goals of the project and questions to be answered are defined and a GQFI table is created following the guidance of the Goal-Question-Metric (GQM) framework [48] with a hierarchical structure. (2) The data on cases (patients), activities and timeframe is extracted from the hospital information system (HIS). (3) Event data is formatted as event logs, and abstracted or filtered as sub-processes, as required for healthcare processes [9]. (4) Event logs are visualized in a log/pattern inspection activity, and a broad overview of the process is obtained in a short time to fully understand the datasets, their limitations and opportunities. (5) After data preprocessing and visualization steps, sets of PM techniques are used in a distributed way by focusing on specific characteristics of the datasets in

terms of questions and indicators. (6) The answers of questions and the distributions of indicators are evaluated, and improvement opportunities are gathered. (7) The findings are used to create proposals for process improvement. It should be noted that many PM techniques and activities are available; however, when dealing with big healthcare data, the executions of PM activities and the tasks of the proposed method should be distributed. Some tasks can be supported with PM tools but not all tasks can be performed using a single PM tool.

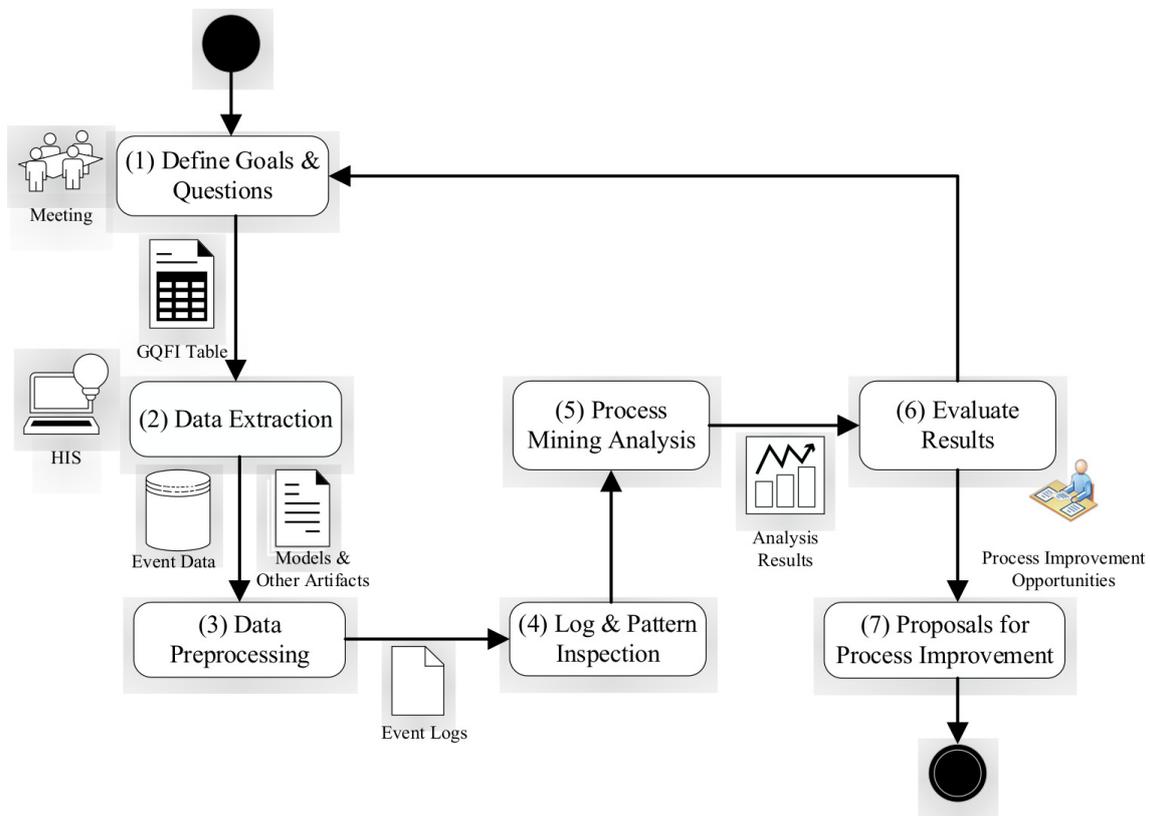


Figure 1. Proposed method.

#### 4.1. Defining Goals and Questions (GQFI)

The aim of this first stage is to exploit domain knowledge from domain experts, end users, customers or management in order to define the project scope, project goals and questions related to these goals. The scope of a project indicates what the process is, when it starts and where it ends, and for which processes and which patients. The goals of a project may be related to improving KPIs related to time, cost, risk, quality, etc. Minimizing mean surgery time, patient waiting time, service response time, and service cost or maximizing service level, i.e., percentage of treated patients within a day, are typical examples. A set of concrete performance-driven questions are used to determine the way to assess or achieve these goals. Each question can be answered with operational scenarios including one or more PM features measuring indicators of the process. Indicators are chosen to provide quantitative answers based on specific needs of stakeholders. PM features, therefore, constitute a bridge between the questions and the indicators, and this bridge can be tailored with respect to the specific needs of stakeholders in PM projects. The output of this stage is the GQFI table.

#### 4.2. Data Extraction

In this stage, patient-related event data is extracted for PM. Healthcare data may be distributed to different data sources that are object-centric rather than process-centric, and may contain incomplete data and outliers at different levels of granularity [37]. Healthcare data attributes are extracted

according to questions and goals as well as the steps of clinical pathways or clinical guidelines. There are three minimum requirements for the control-flow analysis of healthcare data; patient id, information about activities, and timestamps of these activities. Each test, diagnosis, treatment, triage, operation, etc. is a clinical activity. Patient demographics, types and results of tests, name and category of operation, types of triage constitute data attributes for analyzing data enrichment of the discovered process model. Timestamps of data show the sequences and durations of clinical activities and are used for performance analysis. In addition, resource related attributes such as doctor, service, department, emergency or operating rooms are used for organizational analyses.

Existing models are collected for checking the conformance. A hand-made process model (a *de jure* model) is established by interviewing stakeholders, e.g., domain experts, end users, customers, and management. Models, other artifacts (clinical guidelines, pathways or protocols), feedback metrics and KPIs (which are defined and measured by, for example, the quality department of the hospital) are also collected.

The outputs of this stage are the extracted dataset containing event data, data reference models, hand-made process models, feedbacks from experts or stakeholders, KPIs, and other artifacts.

#### 4.3. Data Preprocessing

Mans et al. [3] identified a total of 27 data quality issues (e.g., missing, incorrect, imprecise and irrelevant data). To solve these problems and provide meaningful results, the extracted event data needs to be formatted, filtered, abstracted, or clustered at different levels. Furthermore, event data should be partitioned into pre-and post-mortem event log for further PM analysis.

The event data obtained as a query result must be expressed in the form of event logs to apply PM techniques and connect the event log to the model. This object model consists of the objects of the process, activities, cases, activity instances, events, event attributes (e.g., timestamps, resources, cost and transactions) and the relationships between them at the model, instance and event levels [8].

Questions about a process are answered by filtering, in which the surgery log is given a threshold value, the most-frequent activities are kept, and the events of other activities are filtered out. It can also be helpful to decrease the complexity of the discovered model and reduce the spaghetti effect of the process. When the data is filtered according to a time interval, we can specify a performance target (e.g., 21 days) to determine the cases that do not meet the target value. This feature can be iterative during the PM analysis. The abstraction level is customized by selecting activity or assigning a single value to data attributes. Healthcare data can be converted to different levels using pattern abstraction, e.g., by using a specific plugin in ProM [49].

Clustering can be used to exploratively analyze the event data to gain novel insights into the process. Through clustering from different perspectives (trace clustering, activity clustering, sequence clustering, etc.), the main process can be well identified. Specific characteristics of a dataset may be clustered focusing on the subsets of activities, patients, or followed paths. In this research, the most commonly used techniques were found to be trace clustering, k-means clustering, and hierarchical clustering for healthcare processes [4]. Trace clustering is splitting event logs into homogeneous subsets and creating a process model for each subset. This approach was implemented as a ProM plugin, and is used for complex and diverse event data like healthcare process data [50].

The outputs of the data preprocessing are preprocessed pre-mortem and post-mortem event logs.

#### 4.4. Log and Pattern Inspection

Log and pattern inspection refers to gaining a first impression of the log and gathering statistics to create a log and a pattern summary. This inspection consists of the number of cases, events, duration of events, traces of patterns, resources, frequencies and relative frequencies of events. The aim of this stage is to fully understand and visualize the process in a quick manner. The outputs of this activity can be created in a short time using Disco or ProM plugins.

#### 4.5. PM Analysis and Generating Answers to Questions

Once the data is preprocessed to a suitable format and filtered, different types of PM analysis can be performed. Based on post-mortem and pre-mortem event data and de jure models, PM analysis starts with process discovery techniques adopting a control flow perspective and continues with conformance checking. There are many PM techniques and activities to respond to the questions that arise. This stage includes selecting various PM features, inputs, and outputs obtained within the context of operational scenarios created for each question.

Table 3 lists the PM features and their objectives. After log inspection, the process model is discovered and explored using different mining algorithms. In this stage, the focus is on the discovery of control-flow, organizational aspects (resource perspective), and performance-related information (performance perspective). To visualize the most common paths and outliers, process variants are analyzed. The most complicated or time-consuming cases are visualized or compared in terms of the characteristics of cases. Furthermore, compliance problems and bottlenecks are identified, if any. The cases that do not meet performance targets are visualized. In cross-organizational PM, some analyses are performed using logs from different organizations of the process.

**Table 3.** PM Features and Objectives.

PM Feature	Objective
Process Discovery	Discovering a process model
Process Exploration	Evaluating the discovered process models and exploring a satisfying model
Process Variant Analysis	Discovering process variants (traces, frequencies, and performance measures)
Conformance Checking	Comparing the observed and modeled behavior
Process Enhancement	Extending or repairing existing models using information extracted from the log by replaying the event log on the model
Decision Points Analysis	Extending an existing model with data related attributes (e.g., gender, age, surgery)
Organizational Analysis	Discovering a social network, finding relationships between roles or organizations, or extending an existing model with resource-related attributes
Performance Analysis	Computing performance measures, identifying bottlenecks to trigger a process redesign, or extending an existing model with time-related attributes
Performance Target Analysis	Finding problematic cases by filtering them according to the target value
Process Comparison	Identify how reality deviates from what was planned or expected
Outlier Detection	Identifying outliers to trigger process redesign

The outputs of this stage are discovered models, enhanced process models, compliance problems, bottlenecks, and process variants: e.g., most followed paths, exceptional paths (outliers), most complicated paths, or time-consuming paths.

#### 4.6. Evaluation of the Results

In this stage, the analysis measures of the indicators are evaluated. The inputs are the indicators related to process variants, bottlenecks, and compliance findings from the previous PM analysis stage. After each PM activity, healthcare data is divided into smaller sets. New datasets can be investigated in terms of patient attributes, department, service or doctor information. Examples include analyzing and visualizing the most followed paths to identify the process, analyzing and visualizing the most complicated or time-consuming cases to identify the potential risks, assessing compliance problems for ad-hoc interventions or possible adjustments to the process, and harmonization of two or more process variants to reduce costs or complexity.

#### 4.7. Proposals for Process Improvement

In PM, process improvement is associated with identifying and eliminating bottlenecks, preventing unnecessary states and reprocessing loops and understanding the points that are too time-consuming. Measuring indicators and answering concrete questions provide opportunities for the improvement of the process model. This knowledge can be used to enhance or extend

discovered models. The outputs of this stage are proposals for process improvement and new goals or questions.

## 5. Case Study

The case study was designed and conducted by following the formal case study research approach and guidelines [7,51,52]. The proposed method aims to monitor, control and motivate improvement of healthcare processes and is based on the principles of a goal-driven PM project [8]. The objective of the study is to investigate the applicability of the proposed method and its potential in initiating proposals for healthcare process improvement.

This case study follows the embedded and single-case study design according to the classification proposed by Yin [7]. It is embedded since it considers several analysis units as categories of surgery, and it is a single case study since it considers the surgery process of a university hospital. The clinical data was recorded in January 2017 at all departments of a university hospital in Turkey. 1.152 patients who had undergone an operation were included and the demographics of these patients, surgery type, and the activities related to operation (e.g., admission, starting surgery, finishing surgery, discharge status) were analyzed.

This study uses multiple data sources: archival data as a major source of information, as well as indicators. Archival data is a third degree data type, for which HIS is an important source. The data to be collected is defined based on a goal-driven measurement technique used in the proposed method. In this study, the surgery process of the hospital was selected because it is one of the very well-defined healthcare processes in the organization. The case study was executed by following seven stages of the proposed method and using different PM tools. Four types of PM analysis, namely process discovery, conformance checking, process variant analysis, and performance analysis and two PM tools (ProM and Disco) were used. The latter tool was utilized to generate more understandable and simpler outputs.

### 5.1. Defining Goals and Questions

In this stage, the scope of analysis and the goals, questions, and indicators are determined to decide on the necessary PM features.

In the current case study, the scope of analysis was a general surgery process of a university hospital. Surgery, elective or required, is performed for various reasons. Depending on the diagnosis, the surgery type changes; e.g., optional or elective, required, urgent or emergency. Depending on the type of surgery, different methods, such as open or minimally invasive techniques may be used. According to the American Medical Association and the American College of Surgeons, there are 24 most common surgical operations including appendectomy, cataract surgery, hysteroscopy, mastectomy, etc. The surgery process has several steps; preparation, preoperative tests required, recovery, management of postoperative discomfort and pain management, and discharge planning.

The goal of this study was to understand the timeliness of surgery process from the viewpoint of the management in a university hospital. Four questions were used to determine the way to achieve this goal. Then, a set of indicators which can be measured using PM features were selected. The PM features selected for each question can be answered within relevant operational scenarios. All these elements are presented in a GQFI table as shown in Table 4. This table was created by a PM practitioner after several brainstorming sessions with the healthcare experts in the hospital. More specifically, the indicators were selected from within the KPIs that are subject to audit by the Ministry of Health and after discussions with the quality management staff of the hospital.

**Table 4.** GQFI Table for Surgery Process.

Goal	Purpose	To Understand
	Issue Object (Process) Viewpoint	the timeliness of surgery process from the hospital manager’s viewpoint
Question	PM Feature	Indicators
How does the process look like?	Log and Pattern Inspection	Number of patients/events/activities Average length of hospital stay Start and finish time of process Frequencies of departments/services/doctors/surgeries Number of variants Duration of each patient
	Process Discovery	Identities of process starting/ending activities Number/identities of paths in process map Number of patients/events in each path Number of patients/events in each activity Number of rework loops
What are the most followed paths and the exceptional paths?	Log and Pattern Inspection	Number of variants Number of patients/events in each variant
	Process Variant Analysis	Percentage of patients in each variant Traces (sequence of events) in each variant Identities of patients in each variant Mean duration of each variant Number of outlier patients/events in each variant
Where are the bottlenecks in the process?	Performance Analysis	Total/Mean of duration Total/Mean/Min/Max of duration of each path Total/Mean/Min/Max of duration of each activity Total/Mean/Min/Max of waiting time between activities
Are there deviations from the prescribed process?	Conformance Checking	Matching rate Fitting traces/paths Non-fitting traces/paths Identities of cases in fitting/non-fitting traces

5.2. Data Extraction

After the GQFI table is defined, the data to be analyzed in terms of cases, activities and timeframe is determined. In this process, the event data automatically extracted from HIS with respect to a given query by the data management center of the hospital was used. This data contains information about 1172 cases that underwent any surgical operation in January 2017, for which all diagnostics, treatments and surgical operational activities were recorded.

The surgery process is supported by different departments of the hospital. For any PM analysis of this process, three mandatory parameters are necessary; patient id, activity name, and timestamp. There is a range of different activities undertaken in the history of the patient. There are also additional data attributes which can be used in the analysis; e.g., the doctor that performed the surgery, department, service, operating room, surgery category, diagnosis, and patient demographics such as age and gender.

Contextual data on the processes of ambulatory surgery and inpatient surgery as well as discharge and admission processes was manually obtained from the quality management department of the hospital. The feedback metrics of these processes were as follows: number of cancelled operations due to insufficient preparation before surgery, percentage of outpatients to transferred to a service, length of hospital stay, and number of patients discharged. In addition, mortality rate and number of complications were important indicators showing the efficiency of the surgery process.

### 5.3. Data Preprocessing

This stage consists of formatting and abstraction of sub-processes. Event data should be formatted in event logs to carry out PM techniques. For this reason, the surgery process data was converted automatically and modeled with the software that we developed according to the event object model [8].

Typically, a surgery process comprises the steps of admission to hospital, preoperative examination, surgery, postoperative examination, service transaction, department transaction, intensive care transaction, and discharge activities. A patient can perform any of these activities any number of times. A case object is a patient in this process model, and an event object relates to each activity undertaken for each patient. Age, gender and diagnosis of patient are held at the instance level, and the performed activity, timestamp of activity, department and service of activity, and doctor are held at event level with the event object and its attributes. Age, gender, diagnosis, and surgery name are data attributes of the event; doctor, department, service, and operation room are resource attributes; and the timestamp of activities are the performance attributes. All these attributes can be used to enrich the discovered process model from different perspectives; i.e., data, resource and performance.

Log mining may result in a complicated, spaghetti-like model with too much detail, which makes it difficult to understand and manage data. The surgery log of this hospital contained a huge number of distinct surgical and laboratory activities. To avoid complications, all the low-level laboratory tests were considered as a single test and all distinct operations as a single surgical activity. An example from an event log of a single patient is given in Table 5.

**Table 5.** A fragment of an event log including (for patient 1045) who are diagnosed with atherosclerotic heart disease: event ID, start date, activity, department, and surgery.

Patient ID	Event ID	Activity	Start Date	Department	Surgery
1045	12	Admission to hospital	3 January 2017 09:47	Cardiovascular surgery	Coronary artery bypass
1045	13	Surgery started	4 January 2017 08:00	Cardiovascular surgery	Coronary artery bypass
1045	14	Transfer to ICU	4 January 2017 14:53	Cardiovascular surgery	Coronary artery bypass
1045	15	Surgery finished	4 January 2017 14:55	Cardiovascular surgery	Coronary artery bypass
1045	16	Surgery started	5 January 2017 06:15	Cardiovascular surgery	Postoperative exploration
1045	17	Surgery finished	5 January 2017 07:50	Cardiovascular surgery	Postoperative exploration
1045	18	Transfer to service	13 January 2017 11:20	Cardiovascular surgery	Postoperative exploration
1045	19	Discharged	20 January 2017 11:49	Cardiovascular surgery	Postoperative exploration

### 5.4. Log and Pattern Inspection

Log inspection can be performed using Disco or ProM. ProM is an open source tool that provides many different techniques for log inspection; e.g., dotted chart, log visualizer, log summary plugins, and pattern abstraction. Once the log is loaded, it can be viewed in various ways. Using Disco, the number of events and patients was obtained, the data of individuals was examined from the log, and process variants were traced. Some of the statistics were also checked using the log summary viewer as shown in Table 6 to gain an insight into the surgery process.

**Table 6.** Summary of event logs (as retrieved by Disco).

Indicator	Value
Number of patients	1172
Number of events	6299
Number of activity types	9
Number of surgeries	386
Number of surgery categories	5
Number of departments	25
Number of doctors	97
Number of variants	58
Average length of stay	5 days

### 5.5. PM Analysis and Generating Answers to Questions

In this phase, the results of a detailed analysis of the surgery process are presented. Scenarios including PM features were employed using the existing PM tool of Disco; therefore, in this paper, the outputs (map, statistics and case view) of Disco were used to answer the questions. In the PM analysis stage, each question was answered based on the PM features and the answer was quantitatively evaluated via measuring indicators as follows:

#### Q1 How does the process look like?

The operational scenario to answer this question shows the general characteristics of the process in terms of process map, statistics, and patterns. It includes two PM features explained below.

*Log and Patterns Inspection:* Disco provided detailed statistics of the process. Table 6 shows that there were 1172 cases and 6299 events under nine main activities; thus, the dataset was relatively small. The timeframe of the data was approximately three months. The average length of hospital stay was five days (12.8 days for groups of A1-A2-A3; 5.1 days for group of B; 3.8 days for group of C; 37.8 h for group of D; and 60 h for group of E). For some cases, however, the process lasted up to 129 days. The greatest relative frequency value belonged to the gynecology department (22%), operation room (for day of surgery) (54%), and cataract surgery (7%). There were 58 process variants and five events belonging to the first process variant.

*Process Discovery:* A process map of the surgery process was created to demonstrate how the process looks (Figure 2). This map was generated by Disco Miner, which is a further developed version on Fuzzy Miner [53]. Fuzzy miner is one of the younger process discovery algorithms and uses significance/correlation metrics to interactively simplify the process model at the desired level of abstraction. The starting point of the process map in the figure is a little triangle on the top, which shows 1172 patients. All the patients started either with “Admission to Hospital”, the first step of the process, or directly with the “Surgery Started” activity. After admission to the hospital, the process is split into four alternative paths. The thicker arrow reflects the most frequent activities; e.g., “Surgery Started” activity undertaken by 992 patients after admission. There is no dominant rework loop around the activities; only 25 patients were admitted to the hospital again after having been discharged. Ending activities are “Discharged”, “Ex”, “Surgery Finished”, and “Transfer to a Service”. At the time of the study, 13 patients had not been discharged yet and 1159 patients had completed the surgery process and been discharged. When the activity slider is pulled down to the lowest point, it is seen that there are only five main activities performed. Missing patients may be observed between activities due to the simplified view and the examination of only the most important flows. When the path slider is pulled to the top, all the paths are revealed as shown in Figure 3, and the process map becomes more complicated.

#### Q2 What are the most followed paths and exceptional paths?

The operational scenario for this question includes two PM features explained below:

*Log and Patterns Inspection:* In the case view of this process, there are 58 different process variants belonging to 1172 individual patients.

*Process Variant Analysis:* As shown in Table 7, the most frequent variants of the surgery process were derived. Variant 1 covers around 65% of the events related to 767 different patients (doctors' varying operations and timestamps) that follow the same path within the average length of hospital stay of two days. The most followed path is the same as in the main process. There are many process variants in terms of traces of the surgery process. Twenty-nine variants (0.09%), having a single patient, follow a distinct path from others.

**Table 7.** The most frequent process variants.

ID	Process Variant	Frequency	Percentage(%)
V1	Admission to Hospital > Surgery Started > Surgery Finished > Transfer to Service > Discharged	767	65
V2	Admission to Hospital > Change Service > Surgery Started > Surgery Finished > Transfer to Service > Discharged	89	8
V3	Admission to Hospital > Surgery Started > Surgery Finished > Change Service > Transfer to Service > Discharged	43	4
V4	Admission to Hospital > Surgery Started > Transfer to Intensive Care Unit (ICU) > Surgery Finished > Transfer to Service > Discharged	38	3
V5	Admission to Hospital > Surgery Started > Change Service > Surgery Finished > Transfer to Service > Discharged	32	3
V6	Admission to Hospital > Change Department > Surgery Started > Surgery Finished > Transfer to Service > Discharged	21	2
V7	Admission to Hospital > Surgery Started > Surgery Finished > Transfer to Service > Discharged > Admission to Hospital > Surgery Started > Surgery Finished > Transfer to Service > Discharged	19	2
V8	Surgery Started > Surgery Finished > Transfer to Service > Discharged	17	1
V9	Admission to Hospital > Surgery Started > Surgery Finished > Transfer to ICU > Transfer to Service > Discharged	16	1
V10	Admission to Hospital > Transfer to ICU > Surgery Started > Surgery Finished > Transfer to Service > Discharged	9	1

### Q3 Where are the bottlenecks in the process?

*Performance Analysis:* In the performance analysis using Disco, the total duration and max/min duration showed total/average delays and time spent in different steps of the surgery process, as shown in Figure 4. Following “admission to hospital”, the steps of “change service” (60.9 h) or “change department” (4.2 days) took longer than expected. Furthermore, after being discharged, some patients were admitted to the hospital again within an average of 9.9 days for various reasons. An interesting observation is that some patients waited in the recovery room for 40.3 h on average before being transferred to a service. Figure 4 shows the max and min durations of the paths. The surgery time ranges from 60 secs to 12.2 h. There are some bottlenecks during pre-operative (in transition from “admission to hospital” to “surgery started”) and post-operative (in transition from “surgery finished” to “discharged”) activities of the processes.

### Q4 Are there deviations from the prescribed process?

A detailed view of the process map in Disco tool was used to compare the observed behavior with the modeled behavior. According to the frequency analysis of the whole process (all activities and paths in Figure 3), most cases followed a sequential path: the patient was admitted to the hospital, had surgery, was transferred to a service, and was finally discharged. This conforms to the modeled behavior of the surgery process. Concerning the traces that did not fit the modeled behavior well, the matching rate was calculated as 87% as shown in Table 8. In variants V4, V5, and V15 (n = 75), the “change service” activity appeared to be in the time frame of the patient’s operation. In variants V7 and V13 (n = 25), the patients underwent multiple operations. For six of these patients, the log regarding the transfer to a service was not available. In variants V8, V12, V14, and V16 (n = 37), the “surgery started” activity occurred before “admission to hospital”.

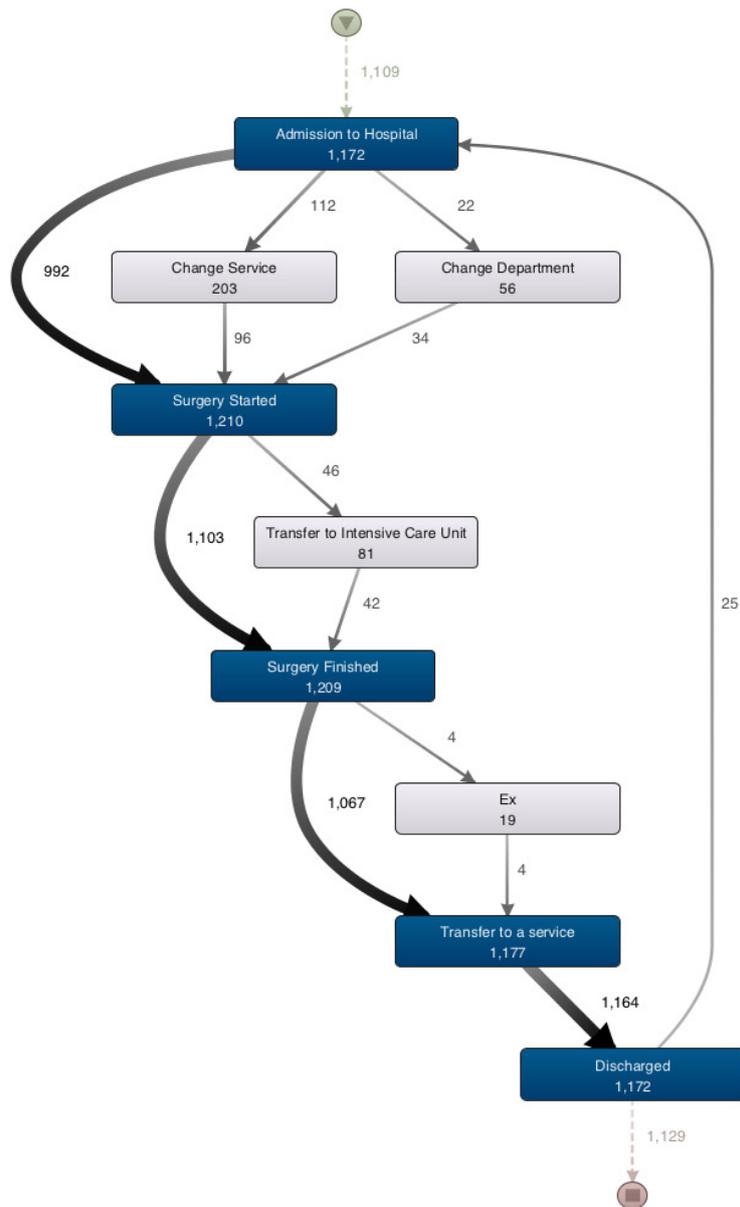


Figure 2. Simplified Process Map of Surgery Process.

Table 8. The most non-fitting process variants.

ID	Process Variant	Frequency	Percentage (%)
V4	Admission to Hospital > Surgery Started > Transfer to ICU > Surgery Finished > Transfer to Service > Discharged	38	3
V5	Admission to Hospital > Surgery Started > Change Service > Surgery Finished > Transfer to Service > Discharged	32	3
V7	Admission to Hospital > Surgery Started > Surgery Finished > Transfer to Service > Discharged > Admission to Hospital > Surgery Started > Surgery Finished > Transfer to Service > Discharged	19	2
V8	Surgery Started > Surgery Finished > Admission to Hospital > Transfer to Service > Discharged	17	1
V12	Surgery Started > Admission to Hospital > Surgery Finished > Transfer to Service > Discharged	9	1
V13	Admission to Hospital > Surgery Started > Surgery Finished > Surgery Started > Surgery Finished > Transfer to Service > Discharged	6	1
V14	Surgery Started > Surgery Finished	6	1
V16	Surgery Started > Admission to Hospital > Change Service > Surgery Finished > Transfer to Service > Discharged	5	0,5

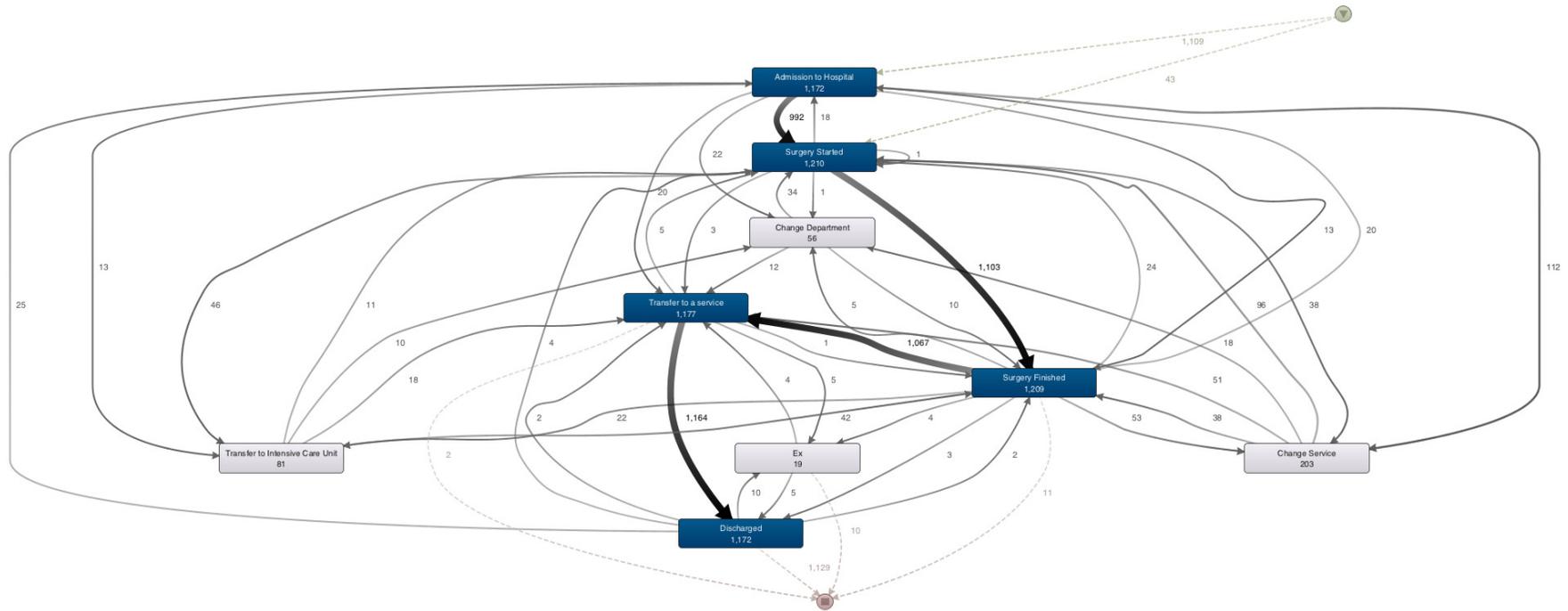


Figure 3. Complete Process Map of Surgery Process.

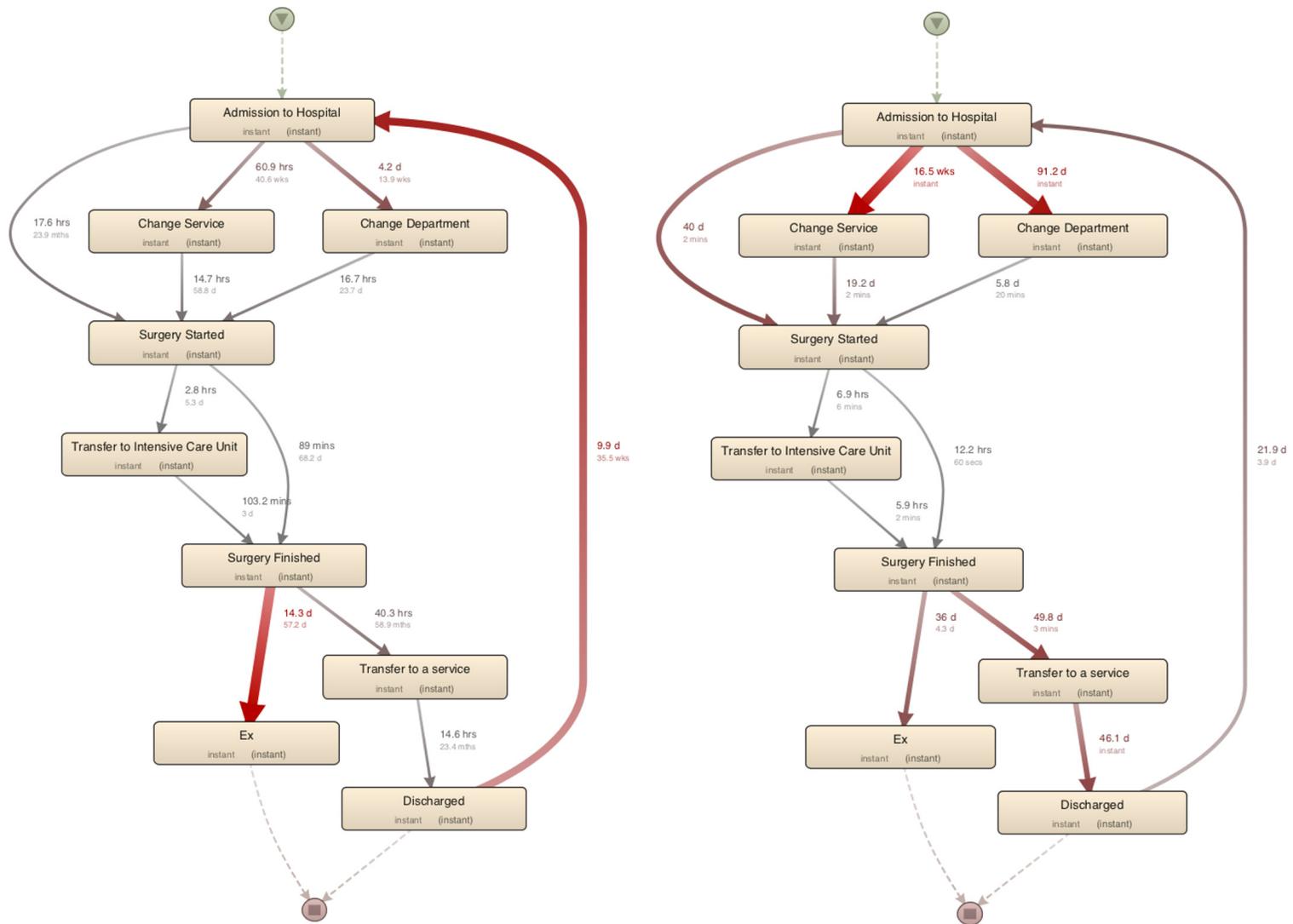


Figure 4. Performance Analysis: (a) Total/Mean Duration (b) Max/Min Duration.

### 5.5.1. Evaluation of the Results

In this stage, an instance of the QOFI table is derived using the actual values of indicators, and the quantitative results are summarized in the form of answers to the questions 1–4.

- Q1 Some performance problems were observed e.g., quite a few cases taking much longer than 21 days (up to 129 days, which is an outlier). It is necessary to determine the step(s) of the process that take(s) too much time, why, and for which patients or surgery.
- Q2 The most followed path covers 65% of all process instances, and there are many exceptional paths (with 29 process variants). These deviations show that the practicing of these processes needs to be divided into different paths for different groups of patients or surgery categories.
- Q3 Bottleneck points of the process are “change service” (60.9 h) and “change department” (4.2 days). This could be due to the problems related to preoperative operations or service resource allocation for these cases. During the average 40.3 h, the patients waited in the recovery room without being transferred to a service. This could be related to the health status of patients or the nature or success of surgery. The process map does not provide information about the reasons for these bottlenecks and the long waiting time since the root causes would differ and need to be investigated.
- Q4 The non-fitting process variants reveal that 6% of the patients were transferred to a service without being officially finished surgery, 3% of the patients had surgery without being officially admitted to the hospital, and six patients were not transferred to a service between two surgical operations. After filtering the non-fitting traces, concrete cases should be investigated with other attributes, such as department, surgery, or surgery category.

### 5.5.2. Proposals for Process Improvement

The surgery process consists of five main activities; admission to hospital, surgery started, surgery finished, transfer to a service, and discharged, which are expected to take place sequentially. Depending on the nature of process or state of patients, between these activities, other activities may also occur (e.g., change service/department or transfer to the intensive care unit or being ex) or loops between them.

In this case study, the timeliness of the surgery process was associated with some basic indicators about the process, variants, deviations, bottlenecks, and inconsistencies. In some of the event logs, skipping of events, missing events, and rework loops (repetitions) of surgery process were detected. There were two bottlenecks resulting from the absence of certain records in pre-operative and post-operative process. Lack of resources and staff or nature of specialty would explain the steps of the process that were delayed. Inconsistencies related to admission and surgery of patient were also observed. Using and integrating RFID tags that record the times for starting and finishing surgery of the patient and visualizing this information could help to increase the quality of care in the hospital.

Finally, new questions were defined based on the bottleneck and inconsistencies identified; e.g., “What are the reasons for deviation from the defined process?” and “Are the records accurate?” These questions allow checking conformance to specifications based on actual behavior, and determine or manage the steps of HIS in which deviations take place. Another question is “How would the process look if we only included patients that stayed at the hospital for more than 21 days?”. Here, we could specify a performance target to examine and visualize the cases that did not meet the target value using the performance target analysis.

## 6. Conclusions and Future Work

In this paper, a goal-driven process evaluation method based on PM was presented for healthcare processes. The method is executed through operational scenarios to answer typical, frequent questions for these processes. The aims, activities, inputs, and outputs of each stage in the proposed method were discussed and then a case study was undertaken to investigate the feasibility of the proposed method in eliciting process improvements. The application of the proposed method covered all three aims of PM activities, as well as other activities such as process variant analysis and performance analysis. A sample GQFI table was derived for the case study, PM activities were executed, the predefined questions were answered, and the results were evaluated. Basic statistics such as mostly followed variants, exceptional variants, bottlenecks, and inconsistencies of the surgery process were analyzed based on PM activities. Opportunities for improvement were identified to minimize cost, reduce average stay time and use of resources, and maximize services and outcomes of the process and the quality of care. Table 9 shows an overview of the SWOT analysis [54] of the proposed method for application of PM in the healthcare domain. Based on the SWOT analysis, it is concluded that strengths and opportunities of the proposed method are more significant than its weaknesses and threats.

**Table 9.** SWOT Analysis.

SWOT Analysis	The Proposed Method
<b>Strengths</b>	Business understanding in terms of goals, questions, and indicators Bridging questions to indicators via PM features Incorporating healthcare domain knowledge to performance evaluation and improvement Tailoring as specific to PM projects
<b>Weakness</b>	Using several tools for distributed PM activities Need for revision of indicators per healthcare specialty
<b>Opportunities</b>	Offering of a comprehensive and iterative methodology Guidance to define operational analyses scenarios Extensibility for new arriving PM features
<b>Threats</b>	Need for revision of features and related indicators as new PM features arrive Need for a supporting tool

In future work, we plan to apply this method to other healthcare processes, e.g., emergency process, by conducting further case studies. In addition, there is a need for the development of a tool that will support the evaluation method and the execution of the operational scenarios and PM features of the proposed method; e.g., process exploration, process comparison, and outlier detection. Developing such a tool might also be useful to support on-site analysis of healthcare process data by medical professionals and hospital managers.

**Author Contributions:** Conceptualization, A.T. and T.G.E.; Methodology, T.G.E.; Validation, T.G.E.; Investigation, T.G.E.; Supervision, A.T.; Writing—Original Draft Preparation, T.G.E.; Writing-Review and Editing, A.T.

**Acknowledgments:** This study received ethics approval from Hacettepe University (approval number GO 18/73, Date: 16 January 2018).

**Conflicts of Interest:** The authors declare no conflict of interest.

## Abbreviations

The following abbreviations are used in this manuscript:

CP	Clinical Pathway
GQFI	Goal-Question-Feature-Indicator
ICU	Intensive Care Unit
PDM	Process Diagnostic Method
PM	Process Mining
PI	Process Improvement
STD	Standard Deviation

## References

1. Kinsman, L.; Rotter, T.; James, E.; Snow, P.; Willis, J. What is a clinical pathway? Development of a definition to inform the debate. *BMC Med.* **2010**, *8*, 31. [[CrossRef](#)] [[PubMed](#)]
2. Kirchmer, M.; Laengle, S.; Masias, V. Transparency-Driven Business Process Management in Healthcare Settings [Leading Edge]. *IEEE Technol. Soc. Mag.* **2013**, *32*, 14–16. [[CrossRef](#)]
3. Mans, R.S.; van der Aalst, W.M.; Vanwersch, R.J. *Process Mining in Healthcare: Evaluating and Exploiting Operational Healthcare Processes*; Springer: Berlin/Heidelberg, Germany, 2015.
4. Erdogan, T.G.; Tarhan, A. Systematic Mapping of Process Mining Studies in Healthcare. *IEEE Access* **2018**. [[CrossRef](#)]
5. Von Alan, R.H.; March, S.T.; Park, J.; Ram, S. Design science in information systems research. *MIS Quart.* **2004**, *28*, 75–105.
6. Hevner, A.; Chatterjee, S. *Design Research in Information Systems: Theory and Practice*; Springer Science & Business Media: Berlin/Heidelberg, Germany, 2010; Volume 22,
7. Yin, R.K. *Case Study Research and Applications: Design and Methods*; Sage publications: Thousand Oaks, CA, USA, 2017.
8. Van der Aalst, W.M. *Process Mining: Data Science in Action*; Springer: Berlin/Heidelberg, Germany, 2016.
9. Rebuge, Á.; Ferreira, D.R. Business process analysis in healthcare environments: A methodology based on process mining. *Inf. Syst.* **2012**, *37*, 99–116. [[CrossRef](#)]
10. Lenz, R.; Reichert, M. IT support for healthcare processes—Premises, challenges, perspectives. *Data Knowl. Eng.* **2007**, *61*, 39–58. [[CrossRef](#)]
11. Masic, I.; Miokovic, M.; Muhamedagic, B. Evidence based medicine—New approaches and challenges. *Acta Inf. Med.* **2008**, *16*, 219. [[CrossRef](#)] [[PubMed](#)]
12. Timmermans, S.; Mauck, A. The promises and pitfalls of evidence-based medicine. *Health Aff.* **2005**, *24*, 18–28. [[CrossRef](#)] [[PubMed](#)]
13. Huang, Z.; Lu, X.; Duan, H. Using recommendation to support adaptive clinical pathways. *J. Med. Syst.* **2012**, *36*, 1849–1860. [[CrossRef](#)] [[PubMed](#)]
14. Sackett, D.L. *Evidence-Based Medicine How to Practice and Teach EBM*; WB Saunders Company: Warsaw, Poland, 1997.
15. Woolf, S.H.; Grol, R.; Hutchinson, A.; Eccles, M.; Grimshaw, J. Potential benefits, limitations, and harms of clinical guidelines. *BMJ* **1999**, *318*, 527–530. [[CrossRef](#)] [[PubMed](#)]
16. Huang, Z.; Lu, X.; Duan, H.; Fan, W. Summarizing clinical pathways from event logs. *J. Biomed. Inf.* **2013**, *46*, 111–127. [[CrossRef](#)] [[PubMed](#)]
17. Hunter, B.; Segrott, J. Re-mapping client journeys and professional identities: A review of the literature on clinical pathways. *Int. J. Nurs. Stud.* **2008**, *45*, 608–625. [[CrossRef](#)] [[PubMed](#)]
18. Cochrane, L.J.; Olson, C.A.; Murray, S.; Dupuis, M.; Tooman, T.; Hayes, S. Gaps between knowing and doing: Understanding and assessing the barriers to optimal health care. *J. Contin. Educ. Health Prof.* **2007**, *27*, 94–102. [[CrossRef](#)] [[PubMed](#)]
19. Rovani, M.; Maggi, F.M.; de Leoni, M.; van der Aalst, W.M. Declarative process mining in healthcare. *Expert Syst. Appl.* **2015**, *42*, 9236–9251. [[CrossRef](#)]
20. Erdoğan, T.; Tarhan, A. Process Mining for Healthcare Process Analytics. In Proceedings of the 2016 Joint Conference of the International Workshop on Software Measurement and the International Conference on Software Process and Product Measurement (IWSM-MENSURA), Berlin, Germany, 5–7 October 2016; pp. 125–130.

21. Van Der Aalst, W.; Adriansyah, A.; De Medeiros, A.K.A.; Arcieri, F.; Baier, T.; Blickle, T.; Bose, J.C.; van den Brand, P.; Brandtjen, R.; Buijs, J.; et al. Process mining manifesto. In *International Conference on Business Process Management, Ferrand, France, 29 August 2011*; Springer: Berlin/Heidelberg, Germany, 2011; pp. 169–194.
22. Scheer, I. *ARIS Process Performance Manager (ARIS PPM): Measure, Analyze And Optimize Your Business Process Performance (Whitepaper)*; IDS Scheer: Saarbruecken, Gemany, 2002.
23. Van der Aalst, W.M.; van Dongen, B.F.; Günther, C.W.; Rozinat, A.; Verbeek, E.; Weijters, T. ProM: The process mining toolkit. *BPM (Demos)* **2009**, 489, 2.
24. Günther, C.W.; Rozinat, A. Disco: Discover Your Processes. *BPM (Demos)* **2012**, 940, 40–44.
25. Rojas, E.; Fernández-Llatas, C.; Traver, V.; Munoz-Gama, J.; Sepúlveda, M.; Herskovic, V.; Capurro, D. PALIA-ER: Bringing Question-Driven Process Mining Closer to the Emergency Room. In *Proceedings of the Business Process Management 2017 (BPM2017)—Demo Session, Barcelona, Spain, 15 October 2017*.
26. Van der Aalst, W.M. Extracting event data from databases to unleash process mining. In *BPM-Driving Innovation in a Digital World*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 105–128.
27. Gatta, R.; Lenkiewicz, J.; Vallati, M.; Rojas, E.; Damiani, A.; Sacchi, L.; De Bari, B.; Dagliati, A.; Fernandez-Llatas, C.; Montesi, M.; et al. pMineR: An Innovative R Library for Performing Process Mining in Medicine. In *Conference on Artificial Intelligence in Medicine in Europe*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 351–355.
28. Fernandez-Llatas, C.; Valdivieso, B.; Traver, V.; Benedi, J.M. Using Process mining for automatic support of clinical pathways design. In *Data Mining in Clinical Medicine*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 79–88.
29. Petersen, K.; Vakkalanka, S.; Kuzniarz, L. Guidelines for conducting systematic mapping studies in software engineering: An update. *Inf. Softw. Technol.* **2015**, 64, 1–18. [[CrossRef](#)]
30. Mariscal, G.; Marban, O.; Fernandez, C. A survey of data mining and knowledge discovery process models and methodologies. *Knowl. Eng. Rev.* **2010**, 25, 137–166. [[CrossRef](#)]
31. Chapman, P.; Clinton, J.; Kerber, R.; Khabaza, T.; Reinartz, T.; Shearer, C.; Wirth, R. *CRISP-DM 1.0 Step-by-Step Data Mining Guide*; SPSS: Chicago, IL, USA, 2000.
32. Azevedo, A.I.R.L.; Santos, M.F. KDD, SEMMA and CRISP-DM: A parallel overview. *IADS-DM* **2008**. Available online: [https://www.researchgate.net/publication/220969845\\_KDD\\_semma\\_and\\_CRISP-DM\\_A\\_parallel\\_overview](https://www.researchgate.net/publication/220969845_KDD_semma_and_CRISP-DM_A_parallel_overview) (accessed on 30 May 2018).
33. Fayyad, U.; Piatetsky-Shapiro, G.; Smyth, P. The KDD process for extracting useful knowledge from volumes of data. *Commun. ACM* **1996**, 39, 27–34. [[CrossRef](#)]
34. Van Eck, M.L.; Lu, X.; Leemans, S.J.; van der Aalst, W.M. PM<sup>2</sup>: A Process Mining Project Methodology. In *International Conference on Advanced Information Systems Engineering*; Springer: Berlin/Heidelberg, Germany, 2015; pp. 297–313.
35. Bozkaya, M.; Gabriels, J.; van der Werf, J.M. Process diagnostics: A method based on process mining. In *Proceedings of the 2009 International Conference on Information, Process, and Knowledge Management, Cancun, Mexico, 1–7 February 2009*; pp. 22–27.
36. Syamsiyah, A.; Bolt, A.; Cheng, L.; Hompes, B.F.; Bose, R.J.C.; van Dongen, B.F.; van der Aalst, W.M. Business process comparison: A methodology and case study. In *International Conference on Business Information Systems*; Springer: Berlin/Heidelberg, Germany, 2017; pp. 253–267.
37. Cardoso, J.; Mans, R.; da Cunha, P.R.; van der Aalst, W.; Berthold, H.; Quaglini, S. A Framework for Next Generation E-Health Systems and Services. *ALSeL* **2015**. Available online: [https://www.researchgate.net/publication/285597858\\_A\\_framework\\_for\\_next\\_generation\\_e-health\\_systems\\_and\\_services](https://www.researchgate.net/publication/285597858_A_framework_for_next_generation_e-health_systems_and_services) (accessed on 30 May 2018).
38. Caron, F.; Vanthienen, J.; Vanhaecht, K.; Van Limbergen, E.; Deweerdt, J.; Baesens, B. A process mining-based investigation of adverse events in care processes. *Health Inf. Manag. J.* **2014**, 43, 16–25. [[CrossRef](#)]
39. Peleg, M.; Soffer, P.; Ghattas, J. Mining process execution and outcomes—Position paper. In *Proceedings of the International Conference on Business Process Management, Brisbane, QLD, Australia, 24–28 September 2007*; Springer: Berlin/Heidelberg, Germany, 2007; pp. 395–400.

40. Fernandez-Llatas, C.; Bayo, J.L.; Martinez-Romero, A.; Benedí, J.M.; Traver, V. Interactive pattern recognition in cardiovascular disease management. A process mining approach. In Proceedings of the 2016 IEEE-EMBS International Conference on Biomedical and Health Informatics (BHI), Las Vegas, NV, USA, 24–27 February 2016; pp. 348–351.
41. Rojas, E.; Sepúlveda, M.; Muñoz-Gama, J.; Capurro, D.; Traver, V.; Fernandez-Llatas, C. Question-driven methodology for analyzing emergency room processes using process mining. *Appl. Sci.* **2017**, *7*, 302. [[CrossRef](#)]
42. Goila, T.; Agrahari, A. A Framework for Evidence Based Business Process Redesign. *SSRN* **2014**. [[CrossRef](#)]
43. Mans, R.; Reijers, H.; Wismeijer, D.; Van Genuchten, M. A Process-oriented Methodology for Evaluating the Impact of IT: A Proposal and an Application in Healthcare. *Inf. Syst.* **2013**, *38*, 1097–1115. [[CrossRef](#)]
44. Cho, M.; Song, M.; Yoo, S. A systematic methodology for outpatient process analysis based on process mining. In *Asia-Pacific Conference on Business Process Management*; Springer: Berlin/Heidelberg, Germany, 2014; pp. 31–42.
45. Yang, S.; Dong, X.; Sun, L.; Zhou, Y.; Farneth, R.A.; Xiong, H.; Burd, R.S.; Marsic, I. A Data-driven Process Recommender Framework. In Proceedings of the 23rd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, Halifax, NS, Canada, 13–17 August 2017; pp. 2111–2120.
46. Dagliati, A.; Sacchi, L.; Zambelli, A.; Tibollo, V.; Pavesi, L.; Holmes, J.H.; Bellazzi, R. Temporal electronic phenotyping by mining careflows of breast cancer patients. *J. Biomed. Inf.* **2017**, *66*, 136–147. [[CrossRef](#)] [[PubMed](#)]
47. 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* **2015**, *15*, 29821–29840. [[CrossRef](#)] [[PubMed](#)]
48. Basili, V.R. *Software Modeling and Measurement: The Goal/Question/Metric Paradigm*; Technical Report; 1992. Available online: <https://drum.lib.umd.edu/handle/1903/7538> (accessed on 30 May 2018).
49. Günther, C.; Rozinat, A.; van der Aalst, W.M.; van Uden, K. *Monitoring Deployed Application Usage with Process Mining*; BPM Center Report BPM-08-11; 2008; pp. 1–8. Available online: <https://pdfs.semanticscholar.org/f677/729e3261b727e2a479b13c17a0f888e8274d.pdf> (accessed on 30 May 2018).
50. Song, M.; Günther, C.W.; Van der Aalst, W.M. Trace clustering in process mining. In Proceedings of the International Conference on Business Process Management, Milano, Italy, 1–4 September 2008; Springer: Berlin/Heidelberg, Germany, 2008; pp. 109–120.
51. Runeson, P.; Höst, M. Guidelines for conducting and reporting case study research in software engineering. *Empir. Softw. Eng.* **2009**, *14*, 131. [[CrossRef](#)]
52. Crowe, S.; Cresswell, K.; Robertson, A.; Huby, G.; Avery, A.; Sheikh, A. The case study approach. *BMC Med. Res. Methodol.* **2011**, *11*, 100. [[CrossRef](#)] [[PubMed](#)]
53. Günther, C.W.; Van Der Aalst, W.M. Fuzzy mining—Adaptive process simplification based on multi-perspective metrics. In Proceedings of the International Conference on Business Process Management, Brisbane, QLD, Australia, 24–28 September 2007; Springer: Berlin/Heidelberg, Germany, 2007; pp. 328–343.
54. Hill, T.; Westbrook, R. SWOT analysis: It’s time for a product recall. *Long Range Plan.* **1997**, *30*, 46–52. [[CrossRef](#)]

