IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Date of publication xxxx 00, 0000, date of current version xxxx 00, 0000.

Digital Object Identifier 10.1109/ACCESS.2024.0429000

PROcess Mining & Simulation for Healthcare Analysis: PROMSHA-Methodology

EVELYN SALAS[1], MICHAEL ARIAS[2], JUAN DIEGO RODRÍGUEZ[3],MACARENA SILVA[4],ARTURO SOLÍS[5], CARLA FERNÁNDEZ[6], SANTIAGO AGUIRRE[7],ERIC ROJAS[8] (Member, IEEE)

[1]School Industrial Engineering, Universidad de Costa Rica, San José, Costa Rica (e-mail: evelynmaria.salas@ucr.ac.cr)
[2]Department of Engineering, Computer Science and Technology, Universidad de Costa Rica San José, Costa Rica (e-mail: michael.arias_c@ucr.ac.cr)
[3]Department of Engineering, Computer Science and Technology, Universidad de Costa Rica San José, Costa Rica (e-mail: juan.rodriguezcambronero@ucr.ac.cr)
[4]Pontificia Universidad Católica de Chile, Santiago, Chile (e-mail: mmsilva8@uc.cl)
[5]Hospital Nacional de Niños, San José, Costa Rica (e-mail: jasolism@ccss.sa.cr)
[6]INCAE Business School, San José, Costa Rica (e-mail: carla.fernandez@incae.edu)
[7]Business Information Systems Department, Central Michigan University, United States (e-mail: aguir2s@cmich.edu)
[8]Institute for Biological and Medical Engineering and Schools of Engineering, Medicine and Biological Sciences and Department of Clinical Laboratories, School of Medicine. Pontificia Unversidad Católica de Chile, Santiago, Chile (e-mail: eric.rojas@uc.cl)

Corresponding author: Evelyn Salas (e-mail: evelynmaria.salas@ucr.ac.cr).

ABSTRACT Process mining (PM) in the healthcare sector is a key discipline in terms of improving processes based on the data stored in clinical information systems. This discipline has been complemented by other techniques, such as process simulation (PS), in order to facilitate decision-making across the sector. In turn, this has enhanced the management of organizational indicators and strengthened the analysis, optimization, and improvement of processes. However, the combination of PM and PS in healthcare still presents a number of challenges, including the incorporation of additional medical specialties, the involvement of healthcare experts, the optimization of data quality, and the promotion of further research that seeks to strengthen the combination of these two areas. The present paper contributes to overcoming these challenges by introducing PROcess Mining & Simulation for Healthcare Analysis: PROMSHA-Methodology, a methodology that provides the necessary steps for process analysis by combining PM and PS in the field of healthcare. The methodology herein addresses certain challenges related to the quality of medical data and includes the participation of experts. To evaluate the usefulness of PROMSHA, a case study was conducted in the pediatric ophthalmology department of a Costa Rican hospital, in which PM techniques were applied to detect spaghetti models and repetitive activities, as well as to identify the causes of waiting lists and bottlenecks by means of several simulated scenarios.

INDEX TERMS Business process management, clinical processes, healthcare, methodology, ophthalmology, pediatrics, process mining, process simulation.

I. INTRODUCTION

THe need to improve response times and resource utilization has prompted the use of analytical approaches to connect operational evidence with healthcare decision-making. The healthcare sector has employed a variety of technologies to support this decision-making, including process mining (PM), which allows organizations to observe how processes are actually executed and identify improvement opportunities using real data [1].

Similarly, PM in healthcare has been complemented by other techniques, such as process simulation (PS), which facilitates the comparison of operational alternatives prior to their implementation in a real-world environment [2], [3], [4]. Moreover, it enables the application of different scenarios to manage bottlenecks, prevent the inadequate use of resources, and optimize process performance [5].

However, literature shows that the joint application of PM and PS in healthcare entails numerous challenges. These include low scientific output, such as different institutional settings or medical areas [6]; how certain clinical specialties have not been studied in depth [7], [8]; that the low quality of

data remains a major issue, and how there are no established methodological frameworks with which to conduct research into these disciplines [6], [9].

Despite these challenges, research into the interplay between PS and PM is considered an emerging field [10], [11], [12], [13], that may help to overcome the limitations of traditional simulation models. Such limitations relate to process timing, interviews with pertinent individuals, and observations for understanding the process. These aspects are traditionally undertaken without the use of computer systems, thus rendering them susceptible to human bias, since the findings are based on perception and a significant amount of time is required to complete the process [10].

Hence, PM helps to identify a process model and enables a simulation model to be developed more quickly than by following this traditional approach [14].

In healthcare, decision-makers are faced with increasingly complex and variable processes on a daily basis. As such, they need more modern methods and techniques to assist in their decision making and to determine the most effective course of action in each case [15], [16], [17]. In this regard, PM and PS can be effective tools. For example, in an emergency department, the use of these disciplines to verify whether additional human resources and/or equipment will significantly contribute to reducing waiting times can completely change the patient pathway and level of care received [18], [19].

The promotion of further research into different health issues through the integration of PM and PS is important as a means of contributing to this decision-making process. However, the absence of structured guidelines with which to facilitate this research and to help increase the number of problems being analyzed hinders the progress of researchers undertaking this work.

Therefore, the present study proposes the PROcess Mining & Simulation for Healthcare Analysis: PROMSHA-Methodology, which serves as a guide for connecting PM with PS in order to analyse healthcare processes by actively involving healthcare experts and highlighting the importance of data quality activities. The methodology was validated in an outpatient consultation process in the ophthalmology department of a specialized pediatric hospital.

The main contributions of the present research include:

1) The introduction of a methodology that applies PM in combination with PS in healthcare.
2) The replicability of the methodology is achieved due to its rigorousness, in terms of its step-by-step approach and graphical visual-based structure, which facilitate overall comprehension thereof.
3) The incorporation of expert validation and data quality management as critical actions.
4) Highlighting the importance of privacy, confidentiality, and data security in healthcare.
5) Serving as a basis to guide healthcare experts in PM and PS analysis, as long as minimum data requirements for such studies are met.

While the present study relies on historical and retrospective clinical event logs, the PROMSHA methodology is not conceptually restricted to static or administrative data sources. Recent research has highlighted the value of continuous and real-time monitoring data, enhancing the accuracy of process mining analyses and simulation models, particularly in healthcare environments, characterized by high variability and dynamic behavior [20], [21], [22]. In this sense, approaches such as the one presented in [20] demonstrate how real-time data streams and continuous monitoring can support more adaptive and precise process representations. PROMSHA can be extended to incorporate such data sources by integrating real-time event acquisition mechanisms into its data extraction and preparation stage, enabling the construction of dynamic process models and more accurate simulation scenarios.

The article is organized as follows: Section II provides information on the main topics included in the proposed methodology; Section III describe the design components for the PROMSHA-methodology; Section IV introduces the methodology and the step-by-step explanation; Section V outlines the results of the methodology validation; Section VI discusses limitations; Section VII presents the discussion; Section VIII addresses conclusions and future work and the Section IX sets out author acknowledgments.

II. BACKGROUND

A. PROCESS MINING AND PROCESS SIMULATION

As a discipline, PM is considered to lie at the intersection of process science and data science. It is based on event logs [1], that are obtained from stored data in information systems [9], [23], and which are used to obtain knowledge about how processes are actually executed. The logs are used to discover models from the data and the verification of compliance with models or guidelines, as well as to propose improvements based on evidence-based feedback [1], [9], [24], [25].

Similarly, PS is a quantitative technique with

which to estimate the performance of a process without interfering with the real environment [26], [27], [28]. It is strategic in contexts of high complexity or uncertainty, when obtaining data or validation in the real environment proves unfeasible or costly, and when analytical solutions are risky [26], [27], [29].

When it comes to programming and designing a process simulation, it is necessary to bring two important concepts to the table: the Digital Twin (DT) and the As-Is Model.

The As-Is Model, or the Process Discovery stage in the Business Process Management (BPM) cycle, aims to reflect the current behavior of the organization regarding how work and the process under study are carried out. These models should represent real processes and assist in identifying improvement opportunities [3].

On the other hand, the DT is also a model of the current process, similar to the As-Is model. It can be seen as a dynamic digital representation of a physical system or process that supports decision-making through real-time data [29], [30], [31], [32]. Its effectiveness lies in integrating the process model under study, as well as elements related to the supply chain, personnel, materials, costs, and all significant information necessary to analyze the real process [13].

When these two techniques are combined, PM provides a solid foundation of actual data, while PS allows the impact of changes to be anticipated, thus enabling experiments to be conducted with different scenarios prior to their real-world implementation [10]. This facilitates the identification of historical behavioural patterns and allows the effects of changes to be predicted prior to their adoption [33]. The integration of PM and PS therefore constitutes a data-driven approach in which event logs feed into the design, configuration of parameters, and validation of simulation models [34], [35], [36].

B. PROCESS MINING AND PROCESS SIMULATION IN HEALTHCARE

PS in healthcare is a high-potential domain due to its contribution to evidence-based improvements. Simultaneously, it faces challenges specific to the sector, including high clinical variability, organizational complexity, barriers to adoption, and data privacy and quality requirements [34], [35], [36], [37].

Additional research into the integration of PM with PS in healthcare is needed [7], [8], [38], [39], since the use of PM for building simulation models is a relatively new field [39]. The literature identifies gaps in this integration, for example:

- A limited application of the two techniques in healthcare institutions [6].
- Medical specialties in which research has been more extensive (cardiology and emergency medicine) compared to others in which no case studies have been identified, such as pediatrics [7], [8].
- Limitations, such as low-quality data, lack of expert validation, and poor integration of clinical sources [8], [23], [35].
- The development of prototypes to automate the generation of models suitable for complex clinical environments [38].
- An absence of solid methodological frameworks, which highlights the need for more structured research to promote their application in real-world contexts [6], [8], [39].

III. METHODOLOGY DESIGN

The PROMSHA methodology was constructed using the Design Science Research method, which is in line with the approach presented by the authors in [40], and is based on the method described in [41]. The main steps of this method are:

1) The Environment: In this phase, we study and analyze the reality of the problem through a systematic literature review. This process allowed us to identify deficiencies in the integration of process mining and process simulation in healthcare, resulting in the lack of a structured guide for conducting studies or research that integrates these disciplines.
2) Knowledge Base: In this second phase, we identified methodologies and methods found in the literature related to case studies, process mining, process mining in healthcare, and process simulation. These methodologies are selected as the knowledge base for constructing the methodology proposed in this work.
3) Design Science Research: In this phase, the methodology is built and applied using the previous findings as input. A graphical structure is defined to facilitate understanding and applicability of the methodology, and the critical methodological factors necessary to combine process management (PM) and process simulation (PS) are identified based on theoretical and practical considerations.

1. The Environment: In 2024, a systematic literature review was conducted, which revealed gaps in research regarding the combination of process mining (PM) and process simulation (PS) in healthcare [6]. One of the main characteristics found in the

primary references of that literature review is that they did not present a structured methodology or guideline that would allow research in these health disciplines to be replicated [6]. The detailed results of this literature review can be consulted in [6]; this gap in the literature has also been identified in other studies [8].

In addition to this systematic review, for the present study, an exploratory analysis was performed using the following keywords: "process mining methodologies and process simulation in healthcare"; "process mining and process simulation in healthcare"; and "methodologies with process mining and process simulation in healthcare. Of the studies reviewed, only five ([38], [42], [43], [44], [45]) outlined an approach that includes a methodology, method, or reference framework. The following noteworthy aspects were identified in the aforementioned studies:

- No visual or graphic structure [45], was identified with which to represent the step-by-step process. In the remaining studies, the proposed approaches did not provide in-depth details of the internal activities of each stage.
- While validation by experts requires greater consideration in general terms, only three studies ([42], [43], [45]) addressed the subject, while none emphasized its importance in a broad manner.
- Despite the use of simulation in [43], its integration into the visual structure of the method presented was not made explicit.
- The critical nature of activities related to the quality of health data was not reflected in any of the studies, nor was the privacy, anonymization, and confidentiality of such data.

Table 1 presents a comprehensive breakdown of the comparison between various factors, including the authors' proposals, the techniques utilized in each method, collaboration with experts, data quality, data privacy and security, and the contribution of the graphical structure. These aspects are examined to highlight their usefulness.

2. Knowledge Base: The study of existing methodologies in areas such as Business Process Management (BPM), PM, PS, and PM in healthcare, served as a reference for structuring the present PROMSHA-methodology (introduced in the subsequent section):

- Case study research ([46], [47])
- BPM lifecycle ([3])
- PM² Methodology ([9])
- FPQ Methodology ([48])

These methodologies were used as a reference due to:

1) The replicability, usability, and scalability of these methodologies
2) The detail in the graphical structure provided
3) The completeness of steps, stages, or activities
4) The ease of understanding for users
5) The in-depth details of each activity of which the methodology or reference framework is comprised
6) These methodologies are frequently cited and used, moreover, because some of their steps are integrated into PROMSHA.

Table 2 presents the details of which phases of each base methodology are being considered in each step of PROMSHA-methodology. Below, an explanation is provided of how PROMSHA relates to each of the reference methodologies:

- Case study research: The proposal in this sense is that the investigations can be developed as case studies; these two basic references stand out in the literature for their usability and replication.
- BPM lifecycle: Since Process Mining serves as a bridge between process science and data science, it is undeniable that it is a discipline that robustly complements Business Process Management (BPM). For this reason, the PROMSHA methodology cannot be detached from the BPM cycle. Includes the stages of process identification, as well as the analysis and redesign thereof.
- PM² methodology: There are several methods in the literature for working with Process Mining; however, we consider that the proposal in the PM² methodology is comprehensive, well-recognized, and robust in terms of phases and ease of application. For these reasons, it is established as one of the foundational methodologies, and all stages of the PM² methodology are covered.
- FPQ methodology: It is necessary to include an existing method that incorporates the healthcare domain; in this case, we draw on the FPQ methodology, as it represents a robust alternative for analyzing healthcare processes using Process Mining. It also integrates a very important component: it strengthens the data extraction activity as a critical stage. All stages are considered within PROMSHA.

3. Design Science Research: Through the application of the methodology in a developed case study, taking an approach based on design methodology

TABLE 1. Comparison of the PROMSHA methodology with existing hybrid approaches.

Methodology, method, framework	Proposal	Techniques used	Collaboration with expert	Data quality	Data privacy and security	Graphic structure
Mesabbah et al. (2019) [38]	Hybrid framework for automatically generating DES models	PM, DES, discovery, extraction	The approach is not explicit	The approach is not explicit	No reference is made	Very simple, it does not provide in-depth details of the internal activities of each stage
Borges et al. (2022) [42]	Framework for ED with PM and simulation	PM, simulation, discovery, variants	It mentions the issue, but does not integrate it throughout the proposal	The approach is not explicit	No reference is made	In general, it does not provide details of the internal activities of each stage
Cho et al. (2014) [43]	Methodology for outpatient analysis based on PM	Process discovery, analysis, what-if, discovery, performance	It mentions the issue, but does not integrate it throughout the proposal	The approach is not explicit	No reference is made	The simulation stage is not even mentioned in the structure
Pegoraro et al. (2018) [44]	Short-term simulation with PM	DES, PM, discovery, performance, discovery	The approach is not explicit	The approach is not explicit	No reference is made	In general, it does not provide details of the internal activities of each stage
Johnson et al. (2019) [45]	ClearPath method with evidence and simulation	PM, simulation, evidence, discovery	It mentions clinical groups and traceability, but does not integrate it across the proposal	It mentions incomplete data, but does not explicitly address the data quality approach	It is partially mentioned	They do not have a structure
PROMSHA-Methodology	To be applied in any clinical setting, where log events can be generated and process simulation can be performed using real data	PM, PS, and remains open to the use of other techniques such as the inclusion of AI. It is not limited to a single type of simulation or PM	The proposal addresses this issue across all activities and includes it as part of the activities to be carried out	It is one of the characteristic aspects of the methodology; it makes it explicit in activities	It is one of the characteristic aspects of the methodology; it makes it explicit in activities	Detailed step-by-step structure with the activities included in each step and additional support defining inputs, outputs, and resources

TABLE 2. Components of each reference methodology included in every step of the PROMSHA-methodology

Steps in PROMSHA	Phases of the reference methodologies integrated into PROMSHA
Step 1.Design	Case study:Planning–Design–Preparation BPM: Process discovery PM²: Planning FPQ: Data extraction
Step 2. Data extraction and preparation	Case study: Data collection BPM: Process discovery PM²: Extraction and data processing FPQ: Data extraction and Event log creation
Step 3.Process mining analysis	Case study: Analysis BPM: Process analysis PM²: Evaluation–Mining–Analysis FPQ: Filtering stage and PM stage
Step 4.Process simulation analysis	No PS methodologies were taken as a baseline
Step 5.Process redesign and improvement	Case study: Share BPM: Process redesign PM²: Process improvement and support FPQ: Results evaluation stage

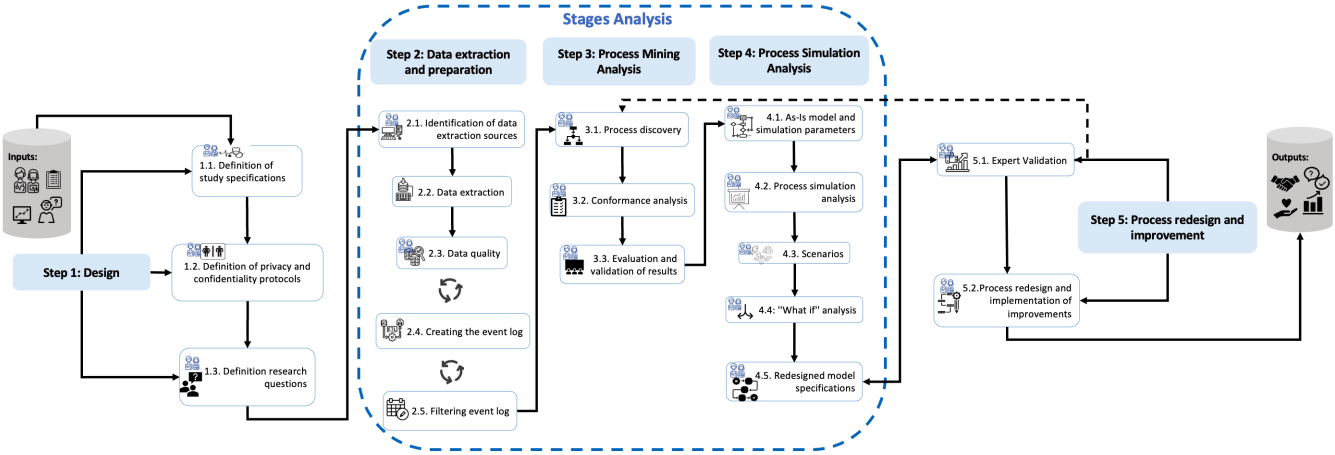

FIGURE 1. Structure of PROMSHA-methodology. The blue people icon represents health experts.

and case study methods [46], [47], three key critical factors that characterize the proposed methodology are identified:

1) Data privacy, security, and confidentiality: emphasizing the importance of protocols for data privacy, security, and confidentiality in healthcare.
2) Data quality: ensuring that the data used in PM and PS analyses are high quality and meet the requirements necessary to answer the relevant research questions.
3) Validation with experts: the involvement of experts is key in order to obtain their feedback and ensure that results are optimal and relevant to clinical practice.

The following section outlines the five steps that constitute the proposed methodology:

IV. PROCESS MINING & SIMULATION FOR HEALTHCARE ANALYSIS:PROMSHA-METHODOLOGY

The methodology proposed by the present study is the PROcess Mining and Simulation for Healthcare Analysis: PROMSHA-methodology. PROMSHA consists of five main steps (see Figure 1): 1. Design, 2. Data extraction and preparation, 3. PM analysis,

4. PS analysis, and 5. Process redesign and improvement.

The central part of the visual structure in Figure 1 integrates steps 2, 3, and 4 as the main analysis phases, while the blue people icon represents the validation link with healthcare experts. The five steps of PROMSHA are explained below:

A. STEP 1: DESIGN

- Activity 1.1. Defining study specifications: The planning, preparation, and design of the study are performed.
  Following meetings with healthcare experts, the medical specialty or specialties under study should be selected; the issues, needs, and expectations of the experts should be defined; and the target population and the period corresponding to the data being used should be established.
  A conceptualization of the environment, process, and medical field is carried out. It may be necessary to interview personnel and experts in the process and justify the problem, to enable the formation of the research team.
- Activity 1.2. Definition of privacy and confidentiality protocols: Ethical protocols related to privacy, anonymization, and data confidentiality are established, as well as guidelines for data extraction and tasks necessary for conducting the research.
- Activity 1.3. Define research questions: The questions that will provide structure to the study are defined, considering expectations and identified issues.

For each step in PROMSHA, a configuration is added with the inputs, outputs, and resources required for each activity. These aspects are not exhaustive; depending on the clinical setting, specialty, expert expectations, and other factors, they may be expanded or varied. Fig. 2 outlines the various aspects of this first step.

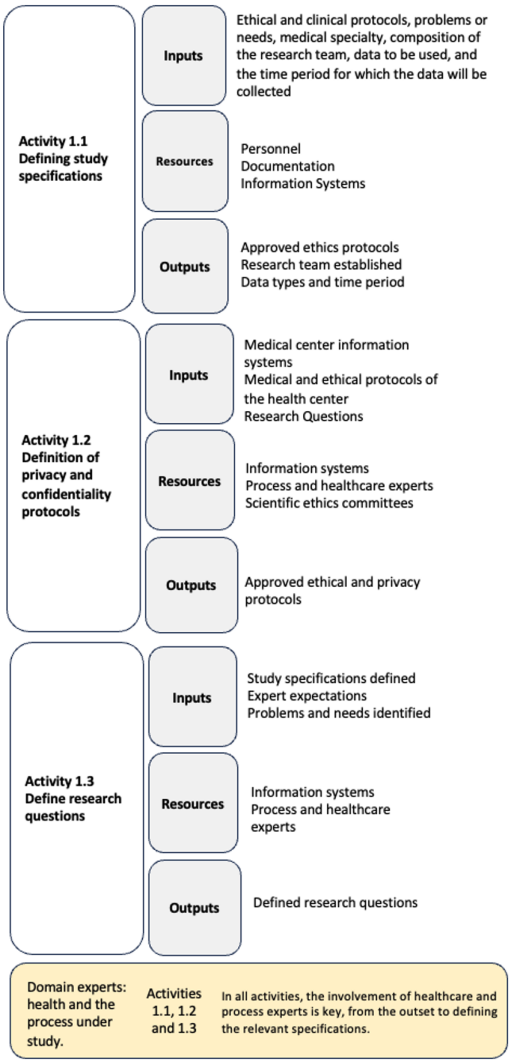

FIGURE 2. Configuration aspects required to execute step 1 of the methodology.

B. STEP 2: DATA EXTRACTION AND PREPARATION

- Activity 2.1. Identification of data extraction sources: The source(s) from which the data will be extracted should be determined. It may be necessary to use more than one source, which is determined in consultation with the personnel responsible for the information systems at the medical center and with healthcare experts.
- Activity 2.2. Data extraction: Extraction can be automated or manual. The source file can be in CSV, XES, JSON, or another format that allows data to be converted to the event log format. The information required for the event log is established, i.e., Case ID, Activity, Timestamp, and any other attributes that need to be extracted.
  In an ideal world, data extraction should be performed automatically, and, as a result, event log generation would also occur automatically. However, not all electronic health records (EHRs) are adapted or process-oriented; therefore, manual extraction must always be included as a relevant stage in this methodology.
  It is important to verify the confidentiality and data handling agreements approved in the protocols, since this will determine who can participate in this process.
- Activity 2.3. Data quality: A cyclical and key activity when working with healthcare data is data quality and cleansing, which can begin at the extraction stage itself. Data quality activities

can be performed in three ways:

1. Non-automated: I.e., people must perform data quality actions. It is recommended that the number of people handling the data be kept to a minimum in order to reduce errors.
2. Automated: I.e., performed using software that allows data quality tasks to be carried out, for example, Power BI[1],Tableau[2] or Qlik[3], Python[4], R[5], SQL[6]. The use of these tools in all research should be declared.
3. Mixed: Combination of non-automated and automated activities.

Any data cleaning or quality issues identified in the dataset must be validated with process and healthcare experts, since they are the ones who should know the actual behavior of the process. If something is incomplete, they must provide the rationale for why it is so. Likewise, the minimum information a case must contain to be considered complete, and a real consultation should be defined with them; otherwise, cases that do not meet the minimum requirements may be removed. Similarly, if there are interruptions in the time series, we must investigate and jointly evaluate to determine whether the remaining data are uncompromised, whether they reflect the characteristics of normal consultations.

The overall quality of the data must be ensured, corresponding to the data quality characteristics[7], such as accuracy, integrity, consistency, conformity, confidentiality, and precision.

- Activity 2.4. Creating the event log: The event log is created according to the research requirements. Each event log may contain additional information about the process, such as the role of healthcare personnel or complementary attributes. This activity can be cyclical, so it may be necessary to perform iterations between the event log and the PM tool to obtain the complete log.
- Activity 2.5. Filtering event log: In the final activity in this step, the event log is prepared using filtering tools to answer the research question(s). In a healthcare setting, for example,

[1]https://www.microsoft.com/es-es/power-platform/products/power-bi

[2]https://https://www.tableau.com/es-es/products/prep

[3]https://www.qlik.com/es-es/products/qlik-sense

[4]https://www.python.org

[5]https://www.r-project.org/about.html

[6]https://www.microsoft.com/es-es/sql-server/sql-server-downloads

[7]https://www.iso25000.com/index.php/normas-iso-25000/iso-25012

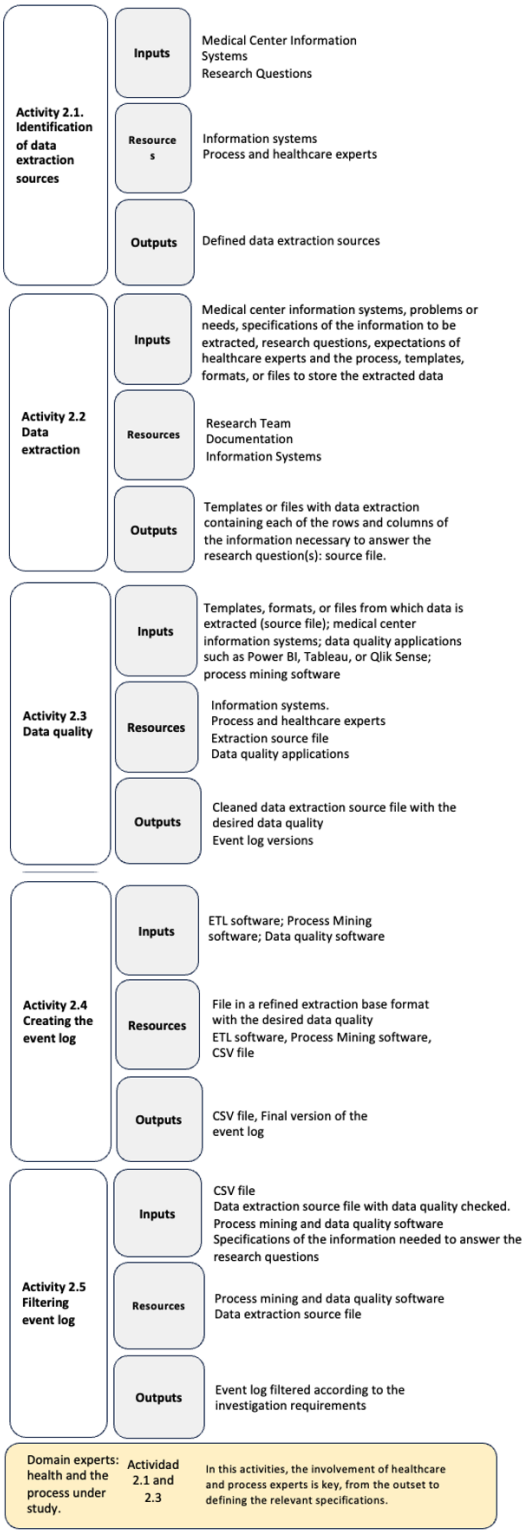

FIGURE 3. Configuration aspects required to execute step 2 of the methodology.

the event log could be filtered according to the desired characteristics, such as the medical specialty under study, the identifier for each case, and specific clinical attributes. There may be several filtering interactions in which different tools are used until the required event log is obtained.

Fig. 3 shows the various aspects of the second step.

C. STEP 3: PROCESS MINING ANALYSIS

Two main analytical perspectives are proposed in the present study: discovery and conformance. However, PROMSHA could be extended to incorporate others, for example, organizational, temporal, and case perspectives.

- Activity 3.1. Process discovery: PM techniques are applied to discover the process in order to obtain a general overview, including different findings about the process behaviour, as well as to answer related research questions.
  This analysis can be performed using one or more tools, such as Celonis[8], Apromore[9], Disco[10] or ProM[11]. Tool selection depends on the research questions, the features of the tool, and the preference of the research team.
- Activity 3.2. Conformance analysis: This activity helps identify inconsistencies between an ideal process model and the log with the actual execution of the process. Research questions related to quality indicators, time, resources, and costs, among others, can be considered.
- Activity 3.3. Evaluation and validation of results: In this activity, the results obtained and the answers to the research questions are analysed. These should be evaluated with healthcare experts in order to validate their importance with regards to the actual process. Such validation is critical because the results form a key part of the simulation analysis (step 4).
  The ability to link PM results to the PS analysis is one of the benefits of the present methodology, since this particular aspect is not addressed in traditional simulations. Therefore, PM acts as a key differentiator. The participation of experts throughout the entire methodology can be addressed by using strategies such as:
- Interviews
- Focus groups
- Brainstorming sessions
- The Delphi method
- Direct interaction during the simulation programming phase
- Active listening for the design of simulation scenarios
- And, for example, interactive presentations of the results from both Process Mining (PM) and Process Simulation (PS).

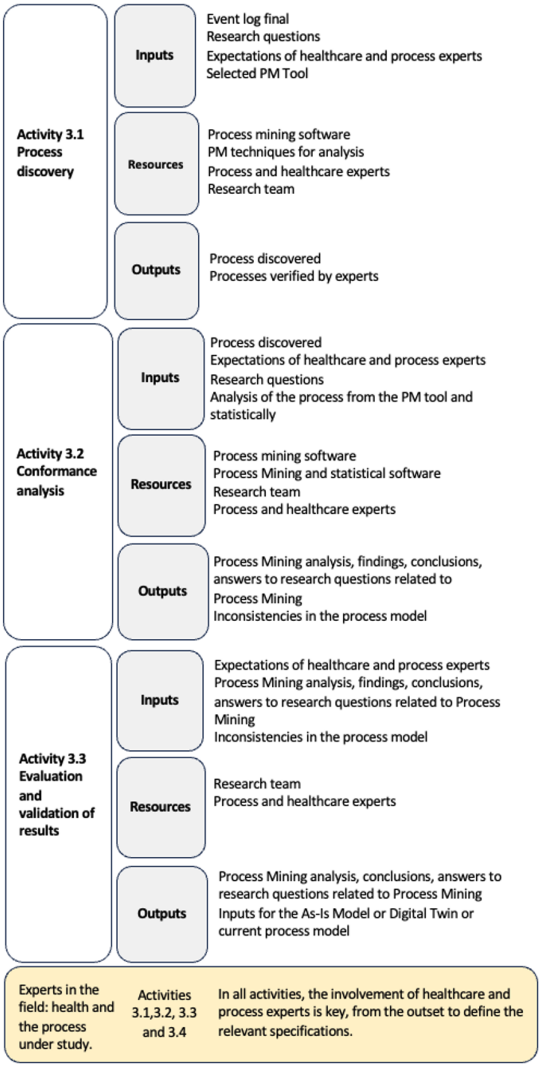

FIGURE 4. Configuration aspects required to execute step 3 of the methodology.

Fig. 4 outlines the various aspects of this third step.

D. STEP 4:PROCESS SIMULATION ANALYSIS

- Activity 4.1. As-Is model and simulation parameters: First, the tool for simulation analysis is selected. If the tool chosen incorporates na-

[8]www.celonis.com
[9]https://apromore.com
[10]https://www.processmining-software.com/tools/disco/
[11]http://www.promtools.org

tive integration of both analyses, the software specifications should be followed in order to begin generating the As-Is model. In this case, the tool should integrate the data from the results obtained in step 3 (PM analysis), while the simulation type predefined by this type of tool should be validated.

It is recommended that for each activity included in the event log, the following be identified: the probability distribution, the standard deviation and the mean. Certain tools may require other statistical information, or information related to the individuals involved in the process, their roles, and weekly work schedules. It is also recommended to consider the allocation of activities performed by each role, the cost per hour, and whether there are any events or particularities in the roles, activities, or the process in general, in order to include these in the As-Is model programming.

If analysis is conducted using software specific to PS, PM input should first be generated and then all relevant aspects of simulation programming addressed, including defining the type of simulation required.

The analysis of the data obtained through PM is used to identify statistical distributions in order to represent the random nature of the data, as well as to establish the level of significance, variance, and the mean. The simulation warm-up process and the simulation period must also be defined, while the number of replicas of the simulation model should be verified to reflect the established confidence level.

Once the simulation programming is complete, the software can be run to generate the results of the current simulated process model. These results can then be validated with domain and process experts in order to verify whether the specifications and results shown correspond to the reality of the process. If not, the necessary adjustments should be made and the relevant programming repeated.

- Activity 4.2. Process simulation analysis: Based on the As-Is model, it is possible to determine the process cycle time, hourly process cost, usability of each role in the process, and execution percentage of the process activities, among other indicators. Running the As-Is model provides useful insights that can be used to help define the scenarios that will be simulated.
- Activity 4.3. Scenarios: The scenarios defined

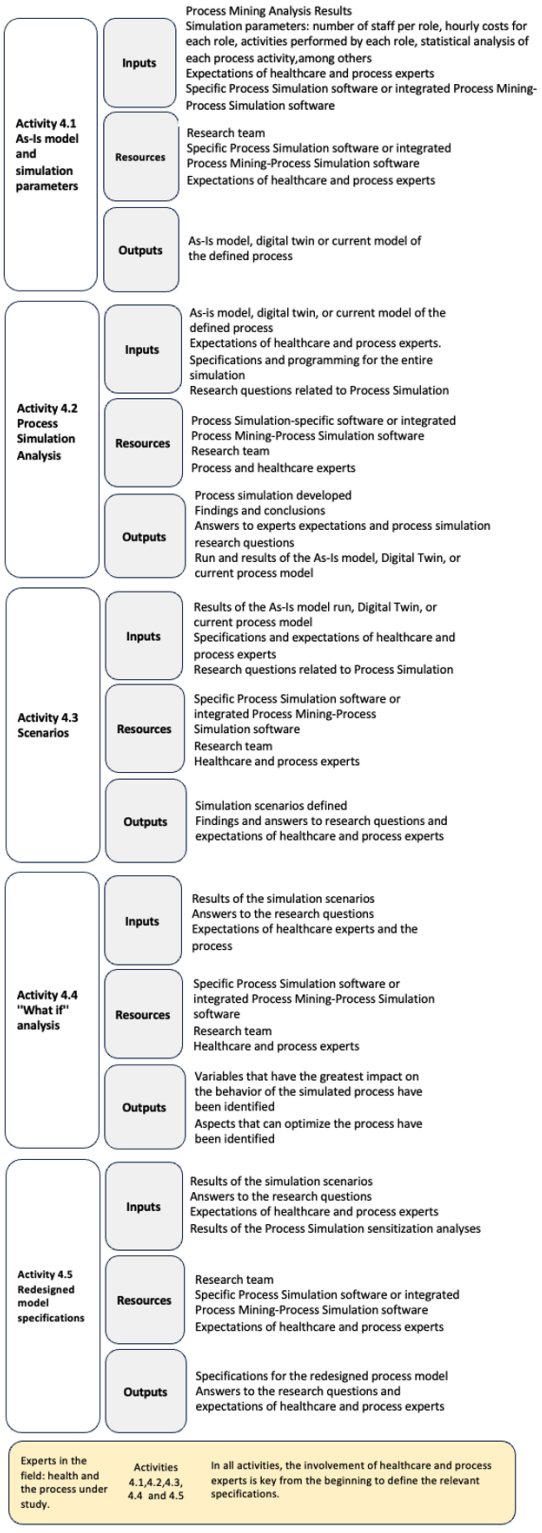

FIGURE 5. Configuration aspects required to execute step 4 of the methodology.

will depend on the expectations of the experts and the analyses proposed as part of the research.

Each scenario can be compared with the As-Is model and the differences between them determined. Furthermore, scenarios can be compared with each other in order to draw conclusions about efficiency and costs related to the simulated changes.

- Activity 4.4."What if" analysis: In this step, the variables with the greatest impact on the behaviour of the simulated process are identified. Consideration is given to what would happen if a variable were changed, adjusted, or modified, and what the impact or consequence of implementing any adjustment would be.
- Activity 4.5. Redesigned model specifications: Based on all the results obtained, including validation and fulfilment of the expectations of the experts, a proposal for improving the process is presented.

A cycle exists between steps 4 and 5. This is due to the fact that once the results are presented to the experts, new requirements for scenarios, assessments, or adjustments may arise and it is necessary to incorporate these into the simulation programming.

Fig. 5 shows the various aspects of the fourth step.

E. STEP 5: PROCESS REDESIGN AND IMPROVEMENT

- Activity 5.1. Expert validation: This activity is critical for the validation of the results as they are managed, as a means to ensure that improvement decisions are not left until the final stage. If any adjustments are required following verification and expert validation, it may be necessary to return to Activity 3.1. Process discovery, in order to redefine new specifications (for example, to include new information in the log, or to adjust the simulation scenarios).
- Activity 5.2. Process redesign and implementation of improvements: The application of the methodology should conclude with a redesign of the analysed process. The aim here is to implement improvements identified during the simulation, which can help to ensure a more accurate understanding of their impact.

Fig. 6 shows the various aspects of the fifth step.

The sequence of steps in the PROMSHA-methodology is not rigid and, as such, it may prove necessary to alternate between different activities or steps. The outcome of each phase determines which phase, or specific activity within that step, should be performed.

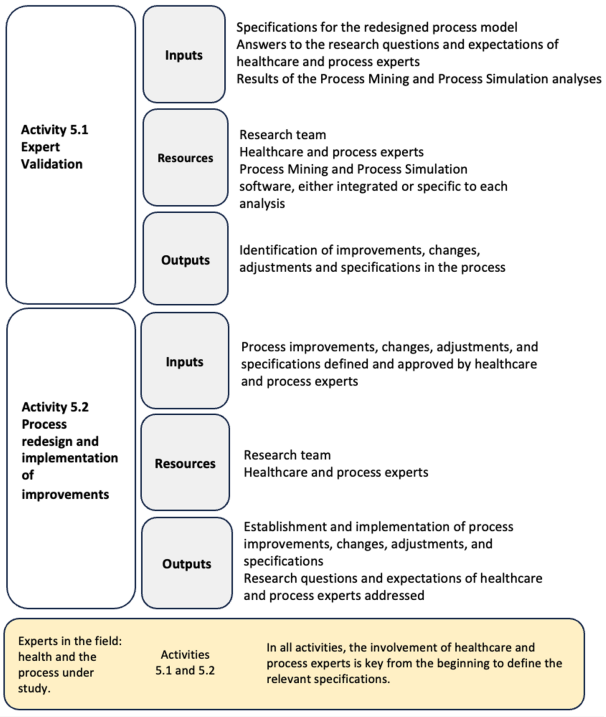

FIGURE 6. Configuration aspects required to execute step 5 of the methodology

V. CASE STUDY

The methodology used in this study followed a case study approach [49], [50], with a focus on reliability as a key quality attribute associated with case studies. "Reliability" refers to the consistency and reproducibility of the findings in a case study [49]. To ensure reliability, it is essential **to document the case procedures through a formal case study protocol** for the tasks to be carried out and to maintain a record of all the steps performed [49], [50]. For this reason, in this case study, each step and activity of the PROMSHA methodology is applied in a detailed and sequential manner.

It is considered that, given that PROMSHA was designed with a high level of detail and by managing the **chain of evidence**—which included inputs, outputs, and resources for each activity within each step, as well as the research questions, the data collected and extracted, the different analyses, and the conclusions—it was possible to develop **data triangulation** using system logs (event logs), expert interviews, and the results of PM and PS analyses. Finally, **expert validation** was applied transversally, involving process managers (For example, the oph-

thalmology coordinator) and domain experts. All these aspects facilitate replicability and enhance the reliability of the proposed methodology.

The methodology was validated by means of a case study conducted at a specialized tertiary-level hospital in Costa Rica, called the Carlos Sáenz Herrera National Children's Hospital (known as the HNN). The tertiary level consists of national and specialized hospitals that handle high-complexity medical and surgical cases and serve as nationwide referral centers for conditions that require specialized technology and skills [51].

The results of the validation at each step of the PROMSHA-methodology are described below.

A. STEP 1: DESIGN

- Activity 1.1. Defining study specifications: The main characteristics of the case study are as follows:
- The present research investigated the outpatient consultation process in a pediatric ophthalmology department. Fig. 7 shows the general activities of this process, as well as the role performed by each activity. The process began with the SOAP (Subjective, Objective, Analysis, Plan) activity. The role that performed the most activities was the doctor, while nursing performed two. Two activities, shaded in grey in Fig. 7, could be performed by both roles.
- Data from January to December 2022 was incorporated, excluding that from May to October, due to a breach in the hospital's information systems during that period.
  October was initially considered for inclusion in the dataset, and its information was extracted. However, after reviewing and agreeing with the head of the outpatient service, and given that October's cases showed an evident lack of information, making them highly incomplete, we decided to remove that month from the dataset.
- Celonis and Apromore were selected as tools due to their ability to perform PM and PS analyses.
- Experts prioritized the study of factors that may be causing waiting lists in the aforementioned specialty.
- Activity 1.2. Definition of privacy and confidentiality protocols: Two ethical protocols were established for data use, privacy, and anonymization. One with the Scientific Ethics Committee of the HNN (CEC-HNN-105-2024) and the other with the Scientific Ethics Committee of the University of Costa Rica (CEC-360-2024).

  These protocols ensure that the data is fully anonymized, eliminating any possibility of tracing an individual's identity. Neither the source extraction file nor the event log includes information such as identity numbers, place of residence, names, surnames, social security numbers, or medical record numbers. Diagnoses are not personalized, nor are any of the extracted activities. Additionally, the protocol stipulates

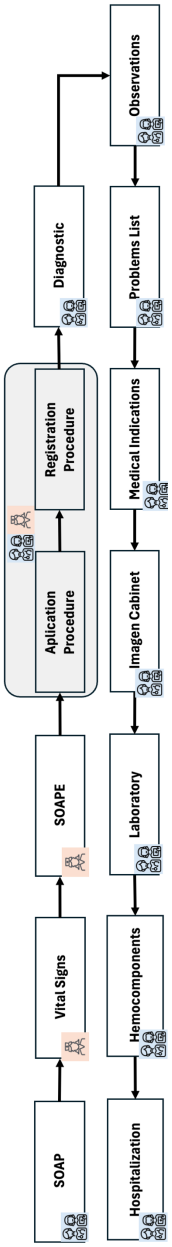

FIGURE 7. Pediatric ophthalmology process activities. The blue people icon represents doctors and the pink one represents nurses.

that data extraction, manipulation, and storage can only be performed by the principal investigator, further reinforcing security measures. Under these protocols, even the counterpart expert involved in the research was not granted access to the data.

- Activity 1.3. Define research questions: In several working sessions with healthcare experts, and considering the problems affecting outpatient services, the following research question was defined:
  Is it possible, through the application of process mining and simulation, to study factors associated with variability in the pediatric ophthalmology care process, for example, those that increase waiting lists, affect process flow, and generate rework?

B. STEP 2: DATA EXTRACTION AND PREPARATION

- Activity 2.1. Identification of data extraction sources: Data was extracted from a single source: an electronic medical record, called the Single Digital Health Record (Expediente Digital Único de Salud) or EDUS.
- Activity 2.2. Data extraction: Data extraction was performed manually from June to November 2024.
- Activity 2.3. Data quality: Multiple automated and non-automated data quality interactions were applied. The Prep Builder application in Tableau software was used to perform thorough data cleansing. These actions took place between November 2024 and March 2025.
  Certain cleansing activities performed included: the standardization of fonts and column sizes; elimination of redundant spaces; identification of hidden or duplicate columns; conversion of numerical data to text; conversion of text fields to dates; and identification and elimination of invalid cells, among others.
  Regarding the missing values, with the support of healthcare experts, it was determined that when a case did not contain the minimum required information (SOAP, diagnosis, start and end time), after applying all data quality procedures, any records that did not accurately represent a consultation were removed from the dataset. This resulted in the removal of 701 rows, leaving a total of 1,814 cases in the event log.
- Activity 2.4. Creating the event log: The file containing the records was converted to CSV format to enable the extract, transform, load (ETL) process. Using the SQL Server Integration Services (SSIS) platform, the information was structured and converted into two event logs.
  One detailed the attributes, while the other recorded the identifier, the activities associated with each case, and the date and time of occurrence. Table 3, shows characteristics of the event log.

TABLE 3. Characteristics of the final event log used for process model discovery.

Characteristic	Value
Number of cases	1814
Variants	385
Number of different activities	27
Roles	3
Violations	27
Specialty	1
Start of log	Jan–2022
End of log	Dec–2022
Average cycle time for each case	61 minutes/case

- Activity 2.5. Filtering the event log: Two event logs were filtered in Celonis, and the decision was taken to use an integrated one in Apromore for tool operability. In addition, the logs were filtered by medical specialty and the identifier for each case was adjusted.

C. STEP 3: PROCESS MINING ANALYSIS

- Activity 3.1. Process discovery: By including the event log in both tools, the actual process flow was identified. For example, Fig. 8 shows the model discovered in Celonis, with 88% of activities and 82% of connections, demonstrating the number of paths actually taken in the process.
  Several final activities were identified, for example, 90% ended following diagnosis, while 7% ended with hospitalization.
  Among the findings, 39% of cases involved a one-hour consultation, which is considerably long since a consultation should last between 15 and 30 minutes.
  Similarly, 6% of cases took between 4 and 10 hours to treat, mainly due to the need to perform some kind of procedure. The aforementioned variability is also reflected in how both extremely short times of 10 to 15 seconds and long times of up to 9 hours were recorded. When this was due to the application of a certain procedure, the expert indicated that this may have

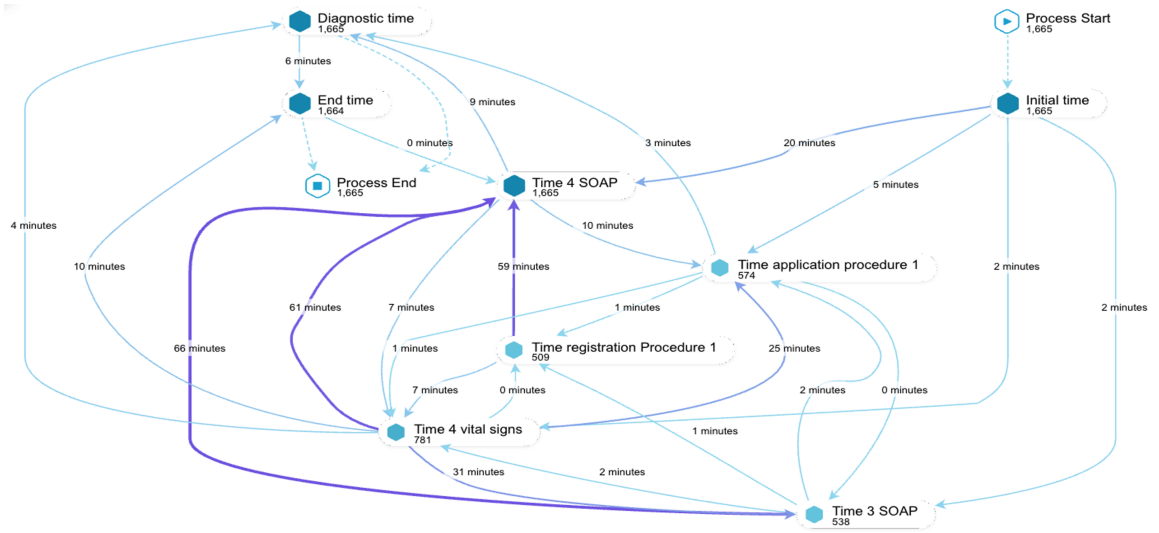

FIGURE 8. **Process model in Celonis**

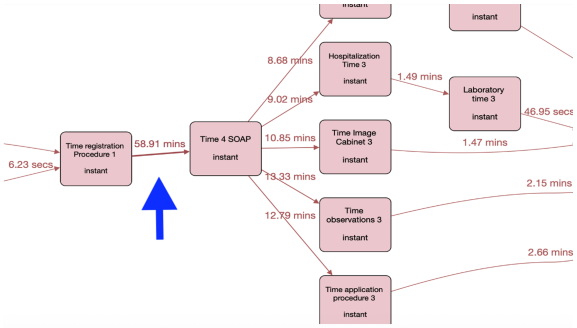

FIGURE 9. **Bottleneck in Apromore. The color of the model is generated by Apromore, representing the view offered by the tool when the duration tab is selected.**

been the result of, for example, the patient not dilating sufficiently and, therefore, medication having to be reapplied and all parties having to wait for the medication to take effect.

- Activity 3.2. Conformance analysis: Findings associated with process compliance included:
  -Only 16.15% of cases followed the most common path, thus it can be assumed that only this percentage of cases complied with the standard. This is the outcome reflected by the data in the Celonis tool; it does not mean that all activities are inefficient. What can be concluded is that this 16% represents the activities that are most frequently repeated in the process, identified as what Celonis calls a "Happy Path," which accounts for that 16%.This conclusion is possible thanks to validation with the experts.
  -The expert stated that Vital Signs monitoring was not necessary, except in specific cases, such as newborns. Nevertheless, it was performed in 48% of cases and added almost 40 minutes to processing time, thereby consuming staff time and failing to add value to patient care.
  -Two instances of reprocessing are evident:
  1. Nurses performed the SOAP activity, a task that nurses are not responsible for executing. This clearly indicates a failure in the use of the system, since this should be carried out in the SOAPE activity (Subjective, Objective, Analysis, Plan, Nursing). Furthermore, in certain cases, the information included in SOAP had already been indicated in the procedure application or registration activity.
  2. When performing the Problems List activity, discrepancies appeared in 44% of cases, thus extending processing time by up to 37 minutes. The information added to the Problems List had already been included in the diagnosis or SOAP activity, which means that the time taken to perform the same activity was doubled. Alternatively, it may reflect misuse of the system.
  -Similarly, three bottlenecks were identified:
  1. When the path from the registration of a procedure to a SOAP activity was available, it was present in 26% of cases and took an average of 59 minutes (See Fig. 9)
  2. If more than one SOAP activity was present, time increased considerably, by up to 66 minutes (41% of cases).
  3. In the path from Vital Signs to SOAP, the increase in processing time was up to 61 minutes (65% of cases). These bottlenecks in Celonis

can also be seen in Fig. 8.

- Activity 3.3. Evaluation and validation of results: All results were validated with the healthcare expert through various working and validation strategies.
  The result validation methods include interview sessions (the questions used are presented in Table 4), result presentation sessions, adjustment incorporation sessions, and cross-checking, involving an additional expert in the presentation of findings and the integration of feedback.
  Table 4 details the strategies, both for methodology design and for validation during the case study.

D. STEP 4: PROCESS SIMULATION ANALYSIS

This step is divided into simulation analysis using Celonis and, subsequently, analysis using Apromore.

In general, for both tools, the cost per consultation is USD155.95 for doctors and USD116.97 for nurses. These costs were provided by the Outpatient Service and Human Resources departments.

Process Simulation Analysis in Celonis:

- Activity 4.1. As-Is model and simulation parameters: Simulation in Celonis begins with the creation of a Digital Twin. Its configuration consists of eight steps:
  1) General: Name of model and currency
  2) Data check: Validation of uploaded data
  3) Filters and granularity: Selection of activities
  4) Incoming cases: Definition of attention schedules
  5) Resources: Assignment of roles and activities
  6) Processing times: Statistical distributions
  7) Pool details: Schedules, costs, and staff numbers
  8) Branches: Sequence probabilities between activities

  A restriction was detected when programming the simulation: Friday could not be included as a working day in the "Incoming cases" step, although it was added in "Pool details," possibly due to low activity recorded on that day, according to the data.
  It was necessary to exclude activities with a small number of cases, since statistical parameters could not be applied thereto (for example, imaging, hemocomponents, and hospitalization).
  When utilizing Celonis, there is no need for a warm-up period for the simulation, unlike with specialized simulation software. These tools allow for the simulation setup to be completed once all necessary information has been entered. After that, the results are generated by clicking the "run" button. Therefore, in this case study, there is no need for preparation time.
  Additionally, the probability distributions are not generated by Celonis or Apromore. Instead, the time data for each of the activities included in the event log had to be processed, and then we decided to use R and Minitab to verify the statistical fits for each activity.
- Activity 4.2. Process simulation analysis: The Digital Twin was configured using the existing process characteristics, which were: 8 doctors and 1 nurse. It was run for 30 days:

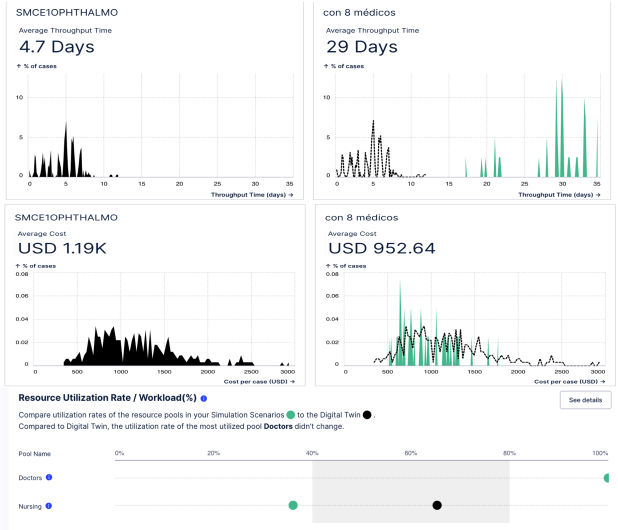

FIGURE 10. Comparison of scenarios 1 and 3 in Celonis

- Average time: 28.4 days
- Average cost: USD977.82
- Resolved cases: 32
- Incomplete cases: up to 298
- Utilization: doctors 100
- Most used activities: SOAP, Problems List, Vital Signs, Diagnostic, Procedure 1
- Most critical activity: Diagnostic (8.62 days)

Following the configuration of the Digital Twin, the simulation was run. Validation with experts confirmed the accuracy of the model, except for the limitation of not being able to include Friday as a working day.

- Activity 4.3. Scenarios: The scenarios were agreed upon with the healthcare expert. Celonis does not allow activities to be deleted directly when programming new scenarios, so the orig-

TABLE 4. Verification and Expert Consultation Strategies

Strategy used to validate with health and in the process experts	**Example of a validated sspect**
Strategies employed in developing the methodology	
Instrument designed to validate critical aspects of the methodology	Validate whether the critical factors defined in the methodology truly facilitate its implementation in healthcare.
Consultation sessions, validation, and discussion of results	The complete methodology was presented to the experts. As a result, they suggested the need to apply it to other specialties within the same outpatient service, in order to compare results and determine whether there are behavioral patterns in the outcomes.
Strategy used to validate with health and in the process experts	**Example of a validated aspect**
Strategies used in applying the methodology (Case Study)	
Collaborative work sessions	From the process discovery analysis using both tools, through the validation of the As-Is model and the Digital Twin, to the findings from the simulation and their implications.
Consultation and validation sessions (structured interviews)	Questions asked to the experts to validate results : 1. Are the cost-related data from the results comparable to the reality of the process? 2. Are the execution time-related data from the results comparable to the reality of the process? 3. Are the resource capacity-related data from the results comparable to the reality of the process? 4. Are the activity occurrence frequency-related data from the results comparable to the reality of the process? 5. Do the activities shown in the Digital Twin, the As-Is model, or the current process model correspond to the reality of the process? 6. What action would you propose for modeling in a simulation scenario? What would you like to see? 7. Is there any indicator, element, or aspect you would like to measure through the simulation? What would you like to achieve?
Cross-validation conducted jointly with both experts in a single session	Present and confirm all results with another expert. This allowed for the identification of additional improvement opportunities and the validation of previously established findings.
Cause-and-Effect Diagram	Identify the possible causes of the problems defined in the PM analyses, which served as inputs for the design of the simulation scenarios.

TABLE 5. Simulated Scenarios and Findings in Celonis

Scenario 1: Optimal Staffing	Scenario 2: Reduced Resources	Scenario 3: Reduced Reprocessing
15 doctors, 1 nurse	8 doctors, 1 nurse	6 doctors, 1 nurse
Removed activities:Problems List, SOAP	Removed activities: Problems List, SOAP	Removed activities: Problems List, SOAP
Remaining activities: Vital Signs	Remaining activities: Vital Signs	Remaining activities: Vital Signs
Throughput time: 4.7 days	Throughput time: 29 days	Throughput time: 28.6 days
Cost per case: $1,190	Cost per case: $952.64	Cost per case: $707.26

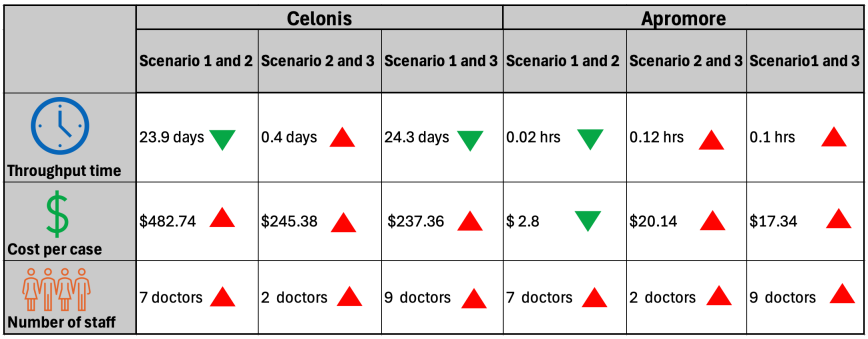

FIGURE 11. Statistical differences between the simulation scenarios in both tools. Upward arrows indicate that the aspect increases, and downward arrows indicate that the aspect has less quantity. For example, in scenarios 1 and 2: scenario 1 has 23.9 fewer days, $482.74 more, and 7 more doctors than scenario 2.

inal Digital Twin was modified to exclude the SOAP, Vital Signs, and Problems List activities, which represented bottlenecks and reprocessing tasks, and different staffing configurations were simulated, as can be seen in Table 5.

When discussing the number of replications in the context of Celonis, it is important to note that there are limitations when generating scenarios. This is because certain parameters, such as the number of activities that can be adjusted, removed, integrated, or automated, cannot be modified within a scenario. Therefore, each scenario had to be created by building a new digital twin to properly address the "what if" questions. Otherwise, only the number of staff, cost, and working hours could be modified. Given this limitation, two runs were performed for each scenario in Celonis.

- Activity 4.4. "What if" analysis: Three key questions were evaluated in the simulated scenarios:

  1. What happens if the number of staff is changed?
  2. What happens if reprocessing activities are removed?
  3. What happens if bottlenecks are addressed?

  The ideal scenario in which a balance is achieved occurs when there are 15 doctors and one nurse.

  The most significant difference between scenarios 2 and 3 is in the cost per case metric, which decreases when the number of doctors is reduced by two. In the case of nursing, the number of staff remains constant, at one, since the usability observed in Celonis does not exceed 66% (see Fig. 10).

  Activities that vary across the three scenarios correspond to reprocessing or those in which discrepancies were identified. The results can be seen in la Table 5.

  In Fig. 11 you can see the statistical differences between the scenarios for each tool, including the differences in staff quantity, cost, and throughput time.

  For example, if scenario 1 were implemented instead of scenario 2, the process duration would be reduced by 23.9 days, which could represent 430 complete appointments currently on the waiting list—assuming each appointment lasts 30 minutes and work is carried out from 7 a.m. to 4 p.m., Monday through Friday.
- Activity 4.5. Redesigned model specifications: The analysis shows that Celonis allows for the exploration of alternatives, the identification of

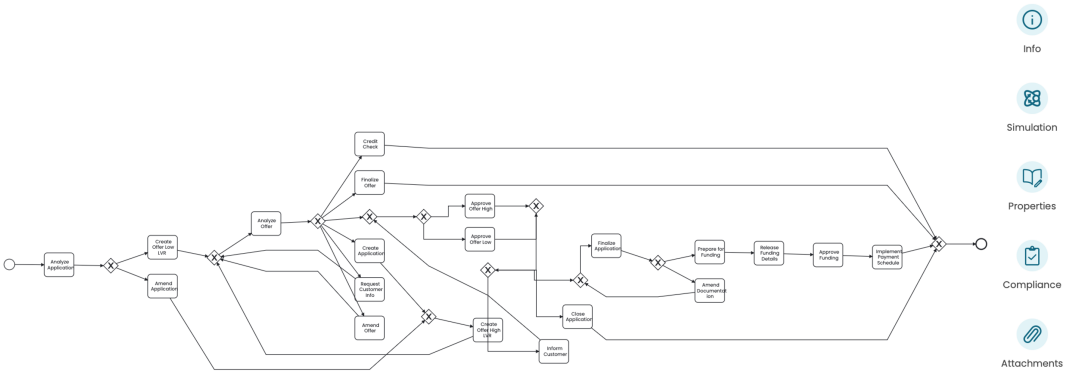

FIGURE 12. **As-Is model in Apromore**

bottlenecks, and the evaluation of changes in resources and reprocessing.

Although it has limitations, this software is useful for decision-making in hospital settings. Celonis was also used to generate process improvements, which were subsequently submitted to the experts.

Proces Simulation Analysis in Apromore:

- Activity 4.1. As-Is model and simulation parameters: Analysis began with the transformation of the process map into a BPMN model, saved as "As-Is" in the "Home folder." This model represented the current state of the process. The experts confirmed its validity.

  When utilizing Apromore, there is no need for a warm-up period for the simulation. These tools allow for the simulation setup to be completed once all necessary information has been entered. After that, in Apromore, the simulation can be initiated by selecting the simulation option. Therefore, in this case study, there is no need for preparation time.
- Activity 4.2. Process simulation analysis: The simulation was programmed in the following nine stages:

1) General: Currency and parameter settings
2) Attributes: Classification of registry data
3) Gateways: Review of process logic
4) Tasks: Statistical data, roles, and costs
5) Functions: Human resources, costs, schedules, and work mode
6) Timetables: Schedule configuration
7) Events: Manual or automatic inclusion
8) Transition times: Times between nodes (optional)
9) Priorities: Conditions based on attributes

The process was then executed in the "simulate model" function, generating a file in the "Home Folder."

- Activity 4.3. Scenarios: The three scenarios were agreed upon with the healthcare expert. The scenarios evaluated efficiency and cost improvements and can be seen in Table 6.

  The As-Is scenario was programmed with the existing characteristics, as follows:
- Staff: 8 doctors, 1 nurse
- Removed activities: Imagen Cabinet;
- Hemocomponents: Doctor instructions
- Cycle time: 1.14 hours
- Cost per case: USD177.08
- Activities per case: 22
- Cases with variations: 13

The configuration of the As-Is model can be seen in Fig. 12.

Regarding the replications, Apromore allowed for the modification of any necessary aspects in the proposed scenarios. Additionally, two runs were conducted for each scenario to ensure a step-by-step alignment with Celonis.

- Activity 4.4. "What if" analysis: The same three questions were explored as in the simulation analysis in Celonis (see Table 6).

  Scenario 1 corresponds to the process configuration in which a balance was achieved. In this case, scenarios 2 and 3 showed a difference of almost 15 minutes in processing time and, as in Celonis, the cost also decreased in scenario 2 compared to scenario 3.

  There is a possibility that the difference in the cost of Celonis' scenarios compared to Apromore's is due to the fact that Apromore was able to count Friday as a working day, and this generates more data for analysis.

TABLE 6. Simulated Scenarios and Findings in Apromore

Scenario 1: Optimal Staffing	Scenario 2: Reduced Resources	Scenario 3: Reduced Reprocessing
Staffing: 15 physicians, 1 nurse	Staffing: 8 physicians, 1 nurse	Staffing: 6 physicians, 1 nurse
Removed: Problems List, SOAP	Removed: Problems List, SOAP	Removed: Problems List, SOAP
Retained: Vital Signs	Retained: Vital Signs	Retained: Vital Signs
Throughput time: 1.13 hrs	Throughput time: 1.15 hrs	Throughput time: 1.03 hrs
Cost per case: $176.24	Cost per case: $179.04	Cost per case: $158.9

Moreover, Tables 5 and 6 show the throughput time in hours for Apromore and in days for Celonis. This is because Celonis allows the simulation run to be defined for a specific time period (in this case, 30 days), whereas Apromore calculates it in hours per case.

In Fig. 11 you can see the statistical differences between the scenarios for each tool, including the differences in staff quantity, cost, and throughput time.

- Activity 4.5. Redesigned model specifications: Apromore can be used to simulate processes, compare scenarios, and make data-driven decisions. Its interface facilitates the identification of improvements, thus optimizing resources and time effectively. Analysis with Apromore helped to generate process improvements that were then presented to the experts.

E. STEP 5: PROCESS REDESIGN AND IMPROVEMENT

- Activity 5.1. Expert validation: Simulation results were shared with experts in collaborative sessions, in which all scenarios, results, and limitations identified were discussed.

  The process of proposing improvement opportunities to the healthcare experts related to the following:

  -- Process standardization
  -- Training staff to use the system
  -- Enable nursing staff to act proactively in procedure application
  -- Remove activities that correspond to reprocessing
  -- Reduce excessive times in procedure application
  -- Improve communication between healthcare personnel involved in the process to adjust response times during procedure application.

  The expert expressed satisfaction with the analysis, highlighting the in-depth nature of the study into the ophthalmological process, which reflected scientific results and was based on real data. The expert also emphasized the need for greater flexibility in the tools, particularly in high-complexity hospital settings, due to the difficulty of including Friday as a working day. This point was determined to be influenced by the scheduling of surgeries, emergencies, and staff vacations.

  The participation of the experts was not only to confirm the results; once the process mining results were verified, the simulation scenarios—the "what if" analyses were defined with them. For example, it was up to the experts to determine when the process is balanced (with 15 physicians and 1 nurse), considering the staffing levels. In addition, it was agreed to eliminate rework and bottlenecks, as well as activities that should not be performed, such as taking vital signs. Likewise, as each step of the simulation was programmed in both Celonis and Apromore, information was validated with the experts—for instance, working hours, hourly costs, and business days. This also helped us understand why Fridays have so few cases and the reasons behind it. These expert decisions are also influenced by the clinical contexts they dominate, which is why they are not only there to validate the results.(See Table 4).

- Activity 5.2. Process redesign and implementation of improvements: In the opinion of the expert, the present methodology enables the study of different organizational indicators, thus offering an objective alternative to subjective perceptions and facilitating evidence-based decision-making.

  At the time of publication of the present study, the head of the outpatient service at the HNN was evaluating the implementation of the proposed improvements and considering extending the use of the PROMSHA-methodology to

other medical specialties.

In general terms, the PROMSHA methodology was fully applicable during the case study; all five of its steps were evaluated, including every activity included within each step. This was made possible thanks to:

- -- Following the PROMSHA methodology step by step.
- -- Meeting the critical aspects: data privacy and security; data quality; and the participation of experts.
- -- The support is provided by identifying inputs, outputs, and resources.
- -- Generating event logs aligned with the research question.

This methodology can be replicated in future studies and differs from other approaches in that:

- -- The level of detail that integrates both the step-by-step narrative and the visual structure, as well as the support provided by inputs, outputs, and resources.
- -- The principle that the generation and filtering of the event log must be carried out in alignment with the research question(s).
- -- The principles of complying with data privacy and security protocols, ensuring data quality, and maintaining the active involvement and participation of healthcare experts throughout the methodology.
- -- The ability to apply PROMSHA not only with retrospective data but also with real-time continuous monitoring data, since it is not limited to a single type.
- -- The openness of PROMSHA to be used in multi-hospital, multi-department, or multi-specialty contexts, as it is not restricted to a single setting; it is possible through the generation of appropriate event logs and the use of suitable PM and PS tools to apply it in those other contexts.

VI. LIMITATIONS

Following validation of the methodology, certain limitations were identified:

- Hospital information systems: Data extraction had to be performed manually since the system did not allow for the extraction of digital records, which resulted in lengthier processing times.
- PM and PS analysis tools: Opportunities for improvement were identified, including making the incorporation of information into simulation programming more flexible, for example, to be able to include Friday as a working day in Celonis; or the need to acquire additional statistical information in Apromore without affecting simulation programming.
- The present study is limited to one particular medical condition, uses data from a single hospital (the HNN), and is limited to a certain number of months in 2022. Consequently, the results cannot necessarily be generalized to other clinical contexts, populations, or time periods. Therefore, it is recommended that future case studies seek to broaden their research scope in order to ensure greater validation.
- The hacking incident suffered by La Caja Costarricense de Seguro Social in 2022 caused: (1) the unavailability of clinical and administrative records for approximately six months, (2) the loss of traceability in critical process events, and (3) temporal gaps that made it impossible to reliably reconstruct certain operational flows. As a result, the data corresponding to that period could not be incorporated to generate the log.
- The current validation of PROMSHA is based on historical and retrospective clinical event records, which represent a limitation of this study; however, PROMSHA is not conceptually restricted to this type of data source. The use of continuous and real-time monitoring data can complement and expand the proposed methodology, enabling future applications of PROMSHA to integrate such data streams. This aims to improve both the quality of event logs for process mining and the accuracy of process simulations.
- The application of this methodology was carried out in a single case study, which may be considered a limitation. Therefore, it is recommended that future research apply it in other clinical specialties, multi-hospital settings, multi-department contexts, or multi–medical-specialty environments. We have several works in progress, where we plan to apply this methodology in other specialties.

VII. DISCUSSION

The PROMSHA-methodology represents a state-of-the-art contribution to healthcare process analysis as it integrates PM and PS in a structured manner. It responds to the need for detailed methodological guidelines in order to help direct studies that combine both disciplines. Its design is based

on proven methodologies in PM, business process management, case studies, and approaches specific to the healthcare sector.

This methodology stands out compared to previous works found in the literature because it identifies the following as critical factors for its application:

1) A clear and detailed visual structure, allowing each activity to be explicitly identified, in addition to all inputs, outputs, and required resources.
2) An iterative approach that promotes close collaboration between healthcare experts, process managers, and researchers.
3) It also highlights the importance of having quality data in order to obtain valid results that accurately reflect reality.
4) Flexibility in data processing, since it can be applied both in environments with "raw" data that require specific cleansing and quality actions, and also in contexts with reliable and structured data.
5) A design that can be applied across different medical specialties and which is easy to use by professionals in distinct areas.
6) And managing privacy, anonymization, and confidentiality of health data.

The application in the pediatric ophthalmology process successfully demonstrated its clinical utility. The results obtained reveal:

- Poor adherence to the ideal process sequence (only 16% of cases follow the ideal sequence), which indicates the need to reinforce standardization and compliance with clinical protocols.
- Identification of bottlenecks, such as when taking vital signs or in the case of there being more than one SOAP activity, since these bottlenecks affect care times and staff efficiency.
- Presence of reprocessing activities, such as the Problems List in the case of doctors and the use of SOAP by nurses, which generate unnecessary workload and operational confusion.
- Inefficient use of certain activities, such as recording Vital Signs with irrelevant populations, which consumes more staff time and affects overall patient care.
- Extreme variability in case duration, ranging from short periods (10–15 seconds) to prolonged ones (up to 9 hours), which complicates planning and resource allocation.

Despite the limitations with the data, this case study made it possible to validate the feasibility and analytical value of the proposed methodology. Just as it allowed us to demonstrate that, through the application of PM and PS, it is possible to study factors associated with variability in the pediatric ophthalmology process, such as bottlenecks, rework, and lack of communication among healthcare staff, which may be impacting the service's waiting lists.

VIII. CONCLUSIONS AND FUTURE WORK

This methodology was validated with a real-life case study conducted in an outpatient service that specializes in pediatric ophthalmology. It was demonstrated that by combining these disciplines health processes can be analysed and decision making supported.

The validation was applied in a case study; however, due to the characteristics of this hospital—being a tertiary-level, high-complexity pediatric facility, in the outpatient service, and located in a country like Costa Rica, where the healthcare system is internationally recognized for its quality—it is considered an appropriate setting to conclude that the results are relevant to the usability of the methodology.

Based on this validation, some recommendations are established to improve the process:

1) Standardization of process execution.
2) Improvement of staff training in the use of systems.
3) Application of segmentation and reordering heuristics to optimize workflows.
4) Elimination of redundant activities.
5) Formalization of effective practices, such as the proactive management of eye dilation by nursing staff..
6) Improvement of communication between doctors and nursing staff to streamline procedure application.

Therefore, the methodology proved to be effective in identifying aspects that impact the efficiency of the process, for example: variability in the process flow (only 16% is compliant); the way the system is currently being used generates rework, bottlenecks, numerous process paths, and unnecessary workload; extreme variations in service times (10 to 15 seconds or 4 to 9 hours) reflect a lack of coordination in the application of procedures. All these findings enable experts and decision-makers to identify improvements in patient care.

The methodology can achieve generalization to multi-specialty or multi-hospital scenarios by generating event logs that align with specific research objectives and questions. It may be necessary to adjust, filter, or normalize the event log, or to use several complementary event logs for comparisons,

for example, but always to ensure that the event logs help answer the research questions and objectives.

By involving healthcare experts, patient-centered results analysis can be conducted. In healthcare, numerous non-operational factors can significantly impact process performance. This is why the involvement of experts is crucial. Example: In the case study where this methodology was applied, it was found that vital sign activity should only be performed for newborns. However, it was observed that this practice is being applied in 48% of cases. This was only possible to identify through consultation with the experts. Therefore, this is one of the strengths of this methodology.

As future work, due to the difficulties with the data used, this methodology is expected to be applied to a complete dataset, whether from other specialties, geographic contexts,in multi-hospital, clinical, or departmental settings or using specific tools in both PM and PS in order to determine its broader usability and application.

Additionally, the use of continuous and real-time monitoring data can complement and expand this methodology, allowing the integration of such data streams into future applications.

It should be noted that a need remains to develop tools that are integrated into hospital systems and which are better suited to the variability and complexity of pertinent services.

IX. ACKNOWLEDGMENT

We would like to thank the Carlos Sáenz Herrera National Children's Hospital of Costa Rica and the outpatient service team for authorizing this case study and for their support in all related matters. We wish to also thank the Universidad de Costa Rica and the Pontificia Universidad Católica de Chile for their support in this research and for the provision of a welcoming academic community in which to operate. Finally, we thank ANID for its support in this research and its ongoing support of Chilean scientists in general. This study forms part of the FONDECYT project #11230708 "Mejorando la Experiencia del Paciente en Telemedicina mediante Minería de Procesos (Improving the Patient Experience in Telemedicine through Process Mining)".

REFERENCES

[1] W. M. van der Aalst y J. Carmona, Process Mining Handbook. Springer Nature, 2022. doi: 10.1007/978-3-031-08848-3.

[2] R. Mans, H. Reijers, D. Wismeijer, and M. Van Genuchten, "A process-oriented methodology for evaluating the impact of IT: A proposal and an application in healthcare," Inf. Syst., vol. 38, no. 8, pp. 1097–1115,2013.

[3] M. Dumas, M. La Rosa, J. Mendling y H. A. Reijers, "Fundamentals of business process management," 2nd ed., Springer, 2018, doi:10.1007/978-3-662-56509-4.

[4] C. Álvarez, E. Rojas, M. Arias, J. Munoz-Gama, M. Sepúlveda y V. Herskovic, "Discovering role interaction models in the Emergency Room using Process Mining," Journal of Biomedical Informatics, vol. 78, pp. 60–77, 2018. doi: 10.1016/j.jbi.2017.12.015.

[5] W. T. de Sousa Junior, J. A. B. Montevechi, R. de C. Miranda, and A. T. Campos, "Discrete simulation-based optimization methods for industrial engineering problems: A systematic literature review," Computers & Industrial Engineering, vol. 128, pp. 526–540, 2019.

[6] E. Salas, M. Arias, S. Aguirre, and E. Rojas, "Combining Process Mining and Process Simulation in Healthcare: A Literature Review," IEEE Access, 2024.

[7] J. Le Lay, J. Neveu, B. Dalmas, and V. Augusto, "Automated generation of patient population for discrete-event simulation using process mining," in Proc. Annu. Modeling and Simulation Conf. (ANNSIM), Jul. 2022, pp. 42–53.

[8] Tamburis, "Bridging the gap between process mining and DES modeling in the healthcare domain," in Proc. E-Health and Bioengineering Conf. (EHB), Nov. 2019, pp. 1–4.

[9] M. L. Van Eck, X. Lu, S. J. Leemans, and W. M. Van Der Aalst, "PM: a process mining project methodology," in Proc. Int. Conf. Advanced Information Systems Engineering, Cham, Switzerland: Springer International Publishing, May 2015, pp. 297–313.

[10] S. Aguirre, C. Parra, and J. Alvarado, "Combination of process mining and simulation techniques for business process redesign: A methodological approach," in SIMPDA 2012, Lecture Notes in Business Information Processing, vol. 162, pp. 24–43, 2013.

[11] H. A. Reijers and W. M. P. van der Aalst, "Short-term simulation: bridging the gap between operational control and strategic decision making," in Proc. IASTED Int. Conf. Modelling and Simulation, pp. 417–421, 1999.

[12] M. T. Wynn, M. Dumas, C. J. Fidge, A. H. ter Hofstede, and W. M. P. van der Aalst, "Business process simulation for operational decision support," in Proc. Int. Conf. Business Process Management, Heidelberg: Springer, pp. 66–77, 2007.

[13] A. Rozinat, R. S. Mans, M. Song, y W. M. P. van der Aalst, "Discovering Simulation Models," Information Systems, vol. 34, pp. 305–327, 2009,doi: 10.1016/j.is.2008.09.002.

[14] G. van Hulzen, N. Martin, and B. Depaire, "The need for interactive data-driven process simulation in healthcare: A case study," in Process Mining Workshops: ICPM 2020 International Workshops, Padua, Italy, October 5–8, 2020, Revised Selected Papers 2, Springer, pp. 317–329, 2021.

[15] H. Eskandari, M. Riyahifard, S. Khosravi, and C. D. Geiger, "Improving the emergency department performance using simulation and MCDM methods," in Proc. Winter Simulation Conf. (WSC), IEEE, pp. 1211–1222, 2011.

[16] L. Vanbrabant, K. Braekers, K. Ramaekers, and I. Van Nieuwenhuyse, "Simulation of emergency department operations: A comprehensive review of KPIs and operational improvements," Computers & Industrial Engineering, vol. 131, pp. 356–381, 2019.

[17] F. Pegoraro, E. A. P. Santos, and E. D. F. R. Loures, "A support framework for decision making in emergency department management," Computers & Industrial Engineering, vol. 146, Art. no. 106477, 2020.

[18] M. Rovani, F. M. Maggi, M. de Leoni, and W. M. P. van der Aalst, "Declarative process mining in healthcare," Expert Systems with Applications, vol. 42, no. 23, pp. 9236–9251, 2015.

[19] W. M. P. van der Aalst, Process Mining – Discovery, Conformance and Enhancement of Business Processes. Heidelberg: Springer, 2011.

[20] U. Celik, A. Korkmaz, and I. Stoyanov, "Integrating Process Mining and Machine Learning for Surgical Workflow Opti-

mization: A Real-World Analysis Using the MOVER EHR Dataset,

[21] C. Imenkamp, A. Maldonado, H. Reiter, M. Werner, W. Hasselbring, A. Koschmider, and A. Burattin, "AVOCADO: The Streaming Process Mining Challenge," arXiv preprint arXiv:2510.17089, 2025.

[22] A. Burattin, "Beamline: A comprehensive toolkit for research and development of streaming process mining," Software Impacts, vol. 17, p. 100551, 2023.

[23] T. Gürgen y A. Tarhan, "Systematic Mapping of Process Mining Studies in Healthcare," IEEE Access, vol. 6, 2018. doi: 10.1109/ACCESS.2018.2831244.

[24] S. Castro y L. Teixeira, "Industry 4.0 and Business Process Management: An Exploratory Study on the Bilateral Effects," Proc. 11th Int. Conf. on Industrial Engineering and Operations Management, Singapore, 2021.

[25] M. Arias, R. Saavedra, M. R. Marques, J. Muñoz-Gama, y M. Sepúlveda, "Human resource allocation in business process management and process mining: A systematic mapping study," Management Decision, 2018.

[26] J. Pereira y P. A. Paulo, "Towards a Characterization of BPM tool's Simulation Support: The Case of BPMN Process Models," International Journal for Quality Research, vol. 13, no. 4, pp. 783–796, 2019, doi:10.24874/IJQR13.04-02.

[27] M.Spacek, P. Pulpán y Z. Kopecky, "Management of business processes using stochastic simulation" SHS Web of Conferences, vol. 83, 2020, Art. no. 01066, doi: 10.1051/shsconf/20208301066.

[28] V. Augusto, X. Xie, M. Prodel, B. Jouaneton, y L. Lamarsalle, "Evaluation of discovered clinical pathways using process mining and joint agent-based discrete-event simulation," en 2016 Winter Simulation Conference (WSC), pp. 2135–2146, IEEE, 2016.

[29] C. de Prada, S. Galán-Casado, J. L. Pitarch, D. Sarabia, A. Galán, y G. Gutiérrez, "Digital Twins in the Process Industry," Revista Iberoamericana de Automática e Informática Industrial, vol. 19, pp. 285–296, 2022. [Online]. https://doi.org/10.4995/riai.2022.16901.

[30] G. D. K. Kuruppu Appuhamilage, M. Hussain, M. Zaman, y W. A. Khan, "A health digital twin framework for discrete event simulation based optimised critical care workflows," npj Digital Medicine, vol. 8, no. 1, p. 376, 2025.

[31] E. Batista and A. Solanas, "Minería de procesos en el ámbito sanitario: Una revisión sistemática," Revista Internacional de Investigación Ambiental y Salud Pública, vol. 18, no. 10, p. 5088, 2021. doi: 10.3390/ijerph18105088.

[32] T. G. Erdogan and A. Tarhan, "Un método de evaluación orientado a objetivos basado en la minería de procesos para procesos sanitarios," Ciencias Aplicadas, vol. 8, no. 6, p. 894, 2018. doi: 10.3390/app8060894.

[33] Günal, M. M., & Pidd, M. (2010). Discrete event simulation for performance modelling in health care: A review of the literature. Journal of Simulation, 4(1), 42-51. https://doi.org/10.1057/jos.2009.25.

[34] E. Rojas et al., "Process mining in healthcare: A literature review," Journal of Biomedical Informatics, vol. 61, pp. 224–236, 2016.

[35] N. Martin, N. Wittig y J. Munoz, "Using Process Mining in Healthcare," en W. M. P. van der Aalst y J. Carmona (Eds.), Process Mining Handbook, pp. 416–444, 2022, doi: 10.1007/978-3-031-08848-3 14.

[36] C. Fernandez, "Chapter 4, Process Mining in Healthcare," en Interactive Process Mining in Healthcare, 1ª ed., ITACA Institute, Universitat Politécnica de Valéncia, Valencia, Spain, Springer, 2021, pp. 41-52.

[37] M. P. Aby, P. Shanmugam, y M. Navonil, "Simulation modelling of hospital outpatient department: a review of the literature and bibliometric analysis," Simulation: Transactions of the Society for Modeling and Simulation International, pp. 1–25, 2022, doi:10.1177/00375497221139282.

[38] M. Mesabbah, W. Abo-Hamad, and S. McKeever, "A hybrid process mining framework for automated simulation modelling for healthcare," in 2019 Winter Simulation Conference (WSC), Dec. 2019, pp. 1094–1102.

[39] W. Abohamad, A. Ramy, and A. Arisha, "A hybrid process-mining approach for simulation modeling," in Proc. Winter Simulation Conf. (WSC), Dec. 2017, pp. 1527–1538.

[40] J. Vásquez, S. Aguirre, C. E. Fuquene-Retamoso, G. Bruno, P. C. Priarone, and L. Settineri, "A conceptual framework for the eco-efficiency assessment of small-and medium-sized enterprises," Journal of Cleaner Production, vol. 237, p. 117660, 2019.

[41] R. Weber, "Design-science research," in Research Methods, Information, Systems and Contexts, pp. 267–288, 2018.

[42] C. M. Borges, S. F. de Sousa, R. O. Coelho, and F. W. Laus, "Framework for Emergency Department Management: A Proposal Based on Process Mining and Simulation."

[43] M. Cho, M. Song, and S. Yoo, "A systematic methodology for outpatient process analysis based on process mining," in Asia-Pacific Conference on Business Process Management, Cham: Springer International Publishing, Jul. 2014, pp. 31–42.

[44] F. Pegoraro, E. A. P. Santos, E. de Freitas Rocha Loures, G. da Silva Dias, L. M. dos Santos, and R. O. Coelho, "Short-term simulation in healthcare management with support of the process mining," in World Conference on Information Systems and Technologies, Cham: Springer International Publishing, Mar. 2018, pp. 724–735.

[45] O. A. Johnson, T. Ba Dhafari, A. Kurniati, F. Fox, and E. Rojas, "The clearpath method for care pathway process mining and simulation," in International Conference on Business Process Management, Cham: Springer International Publishing, Sep. 2018, pp. 239–250.

[46] K. M. Eisenhardt and M. E. Graebner, "Theory building from cases: Opportunities and challenges," Academy of Management Journal, vol. 50, no. 1, pp. 25-32, 2007.

[47] R. K. Yin, Case study research: Design and methods, vol. 5. Thousand Oaks, CA: Sage, 2009.

[48] E. Rojas, M. Sepúlveda, J. Munoz-Gama, D. Capurro, V. Traver, and C. Fernandez-Llatas, "Question-driven methodology for analyzing emergency room processes using process mining," Applied Sciences, vol. 7, no. 3, p. 302, 2017.

[49] Teegavarapu, J. D. Summers, and G. M. Mocko, "Case study method for design research: A justification," in Proc. Int. Design Engineering Technical Conf. and Computers and Information in Engineering Conf., vol. 43284, pp. 495–503, Jan. 2008.

[50] E. Yacuzzi, "El estudio de caso como metodología de investigación: teoría, mecanismos causales, validación," Serie Documentos de Trabajo, no. 296, 2005.

[51] M.D.R. Sáenz, M.Acosta, J. Muiser, & J.L.Bermúdez, "Sistema de salud de Costa Rica". Salud pública de méxico, 53, s156-s167, 2011.

EVELYN SALAS Received the master's degree in industrial engineering from the Universidad de Costa Rica. She is currently pursuing the Ph.D. degree in process mining and process simulation. Her professional focus is on the area of organizational development, process management, and process simulation. She is currently a professor of industrial engineering with the Universidad de Costa Rica. Currently completing his doctoral internship at the Institute of Medical and Biological Engineering (IIBM) of the Faculty of Engineering, Medicine, and Biological Sciences at the Pontificia Universidad Católica de Chile.

MICHAEL ARIAS Received the Ph.D. degree in engineering from the Pontificia Universidad Católica de Chile (PUC), with a focus on computing. He is a Full Professor and a Researcher with the Universidad de Costa Rica (UCR). He is currently with the Human and Process Research Laboratory (HAPLAB), Department of Computer Science (DCC), PUC, and leads the Celonis Process Mining Center of Excellence, UCR. His research interests include process mining, business process management, process-oriented data science, project management, human resource management, IT governance, and gender in computer science.

JUAN DIEGO RODRÍGUEZ Master's Degree in Computing and Informatics from the University of Costa Rica. Since 2009, works at CODISA (now Novus ERP) as Infrastructure Coordinator. Has developed Business Intelligence (BI) and Data Analysis projects in markets such as banking, industry, and services. Manages everything related to O365 tenants and Oracle Cloud Infrastructure. Has been teaching at the University of Costa Rica since 2018, focusing on training in topics related to teaching, business process mining, artificial intelligence, programming, and networking. A researcher passionate about process mining, actively participates in the Celonis Center of Excellence, combining process intelligence with business intelligence.

MACARENA SILVA Is currently pursuing the B.S. degree in Biomedical Engineering, with a minor in Information Technologies, at Pontificia Universidad Católica de Chile (Santiago, Chile). Her interests include healthcare process mining, patient safety and quality, and clinical performance analytics.

ARTURO SOLÍS Born in San José, Costa Rica. Graduated in Medicine & Surgery at the University of Costa Rica (UCR) in 1993. After working as a general practitioner for 2 years, he then pursued a specialization in Paediatrics, also at UCR. In 1998, he moved to Liverpool, UK, where he worked as a clinical fellow of paediatric respiratory medicine at Alder Hey Children's Hospital for four years, getting involved in both clinical and research duties. He currently works at the Hospital Nacional de Niños in Costa Rica, a tertiary paediatric hospital with over 300 beds. He is the author of several scientific articles, an active reviewer on the Cochrane Library, and an author and editor on Frontiers in Paediatrics. Dr. Solís Moya is professor of paediatrics at the University of Medical Sciences and a member of the Costa Rican Association of Pulmonology and Thoracic Surgery.

CARLA FERNÁNDEZ Professor Carla Fernández Corrales earned her Ph.D. in Business Administration with a concentration in Information Systems from the University of Illinois at Urbana-Champaign. She completed undergraduate studies in Industrial Engineering and graduated with honors, receiving a Master of Science degree in Computer Science, both from the University of Costa Rica. She was honored with the Fulbright scholarship from the United States Department of State. Dr. Fernández Corrales' research focuses on technology management and innovation understanding through machine learning methods. Her teaching interests center around the strategic use of emerging technologies, data science, and agile methods. As a consultant, speaker, and TEDx speaker, Professor Fernández Corrales is distinguished for her ability to convey highly specialized technological knowledge to non-technical audiences, establishing connections with value propositions, strategy, and risk analysis in organizations. Passionate about innovation and the digital transformation of higher education, she has also held various positions of academic leadership. These include serving as the Director of Educational Technologies (METICS) at the University of Costa Rica, the Director of Learning Innovation at INCAE Executive Education, and the Academic Director of MBA programs at INCAE Business School. Continuing her mission of transforming higher education in times of disruption, she currently serves as Senior Director of Executive Education and INCAE Online at INCAE Business School.

SANTIAGO AGUIRRE Received the Ph.D. degree in engineering from Pontificia Universidad Javeriana, Colombia. He has over two decades of teaching and research experience in fields related to process mining, information systems, data analytics, process automation, and business process management (BPM). He is currently a Faculty Member with the Business Information Systems Department, Central Michigan University, USA. He previously held the position of a Professor with the Industrial Engineering Department, Pontificia Universidad Javeriana. He has published record 14 peer-reviewed articles, 16 conference papers, and one book.

ERIC ROJAS Received the Ph.D. degree in engineering from the Pontificia Universidad Católica de Chile (PUC), with a focus on computing. His research interests include clinical informatics since his Ph.D. thesis in process mining in healthcare, in 2017. He is currently an academic with the Department of Clinical Laboratories, Faculty of Medicine, PUC, and the Leader of the Quality Assurance Area, Chilean National Center for Health Information Systems (CENS). In addition to being a member of the Institute of Biological and Medical Engineering, he is also an academic at the Department of Clinical Laboratories of the PUC Faculty of Medicine and Leader of the Quality Area at the National Center for Health Information Systems.