Process Mining in Healthcare: A Study About Data Standardizing Framework, KPIs, and Domain Expertise Involvement in Emergency Care of Developed and Non-Developed Countries

1st Krupa Srinivas
Master's in Embedded Systems Engineering, *University of Twente*
Enschede, The Netherlands
krupasrinivas@student.utwente.nl

***Abstract*—Emergency departments (EDs) play a critical role in the healthcare system but face challenges in patient outcomes and operational efficiency. The absence of a standardized framework for data quality and key performance indicators (KPIs) prevents the identification of inefficiencies and benchmarking of ED performance. Additionally, process mining insights often lack validation from domain expertise, limiting their practical applicability to real healthcare settings. This research adopts the CRISP-DM framework to address these challenges using the MIMIC-IV ED datasets. The study focuses on improving data quality using the Observational Medical Outcomes Partnership (OMOP) Common Data Model (CDM), standardizing KPIs for performance evaluation and integrating domain expertise to validate process mining insights across developed, underdeveloped and developing countries. The study follows a structured Extract-Transform-Load(ETL) process to prepare and standardize ED data for process mining. Process discovery techniques are applied to derive KPIs such as rework rate, average processing time, case resolution rate, and patient throughput. Process discovery is performed using a power automate tool by Microsoft. Trend analysis to further explore seasonal impacts, arrival processes, chief complaints and patient disposition. Domain expertise is integrated to validate these findings and align them with clinical practices globally. This study contributes to improving operational efficiency, optimizing workflows and enhancing patient outcomes in emergency departments worldwide.**

***Index Terms*—Healthcare, data standardisation, data quality, domain expertise, MIMIC IV ED dataset, ETL, key performance indicators, OMOP Common Data Model (CDM), patient outcomes**

I. Introduction

In the post pandemic world, the healthcare industry has recognized the importance of optimizing operational efficiency and improving patient outcomes. Focusing on Emergency Departments (EDs) that has dynamic and complex processes, several challenges require attention and improvements. This has led to potential research and implementation of process mining techniques in the healthcare industry. EDs are a critical area where process mining can significantly yield benefits as they require effective management to ensure patient safety and satisfaction. The key challenges identified in this research are data quality and integration, the assessment of performance metrics and the essential involvement of domain expertise to ensure processes are accurate, actionable and reliable. High-quality data and robust performance metrics are critical to accurately evaluate healthcare processes in emergency departments [3]. However, process mining in healthcare is fundamentally challenged by the nature of data generation within healthcare system [36]. Data within these systems is often incomplete, erroneous and fragmented due to varied information systems for data collection. Patient data is typically distributed across multiple systems, such as electronic health records (EHRs), laboratory database, radiology and other specialized departments, each with its own data format and standards. This fragementation leads to inconsistencies in the data and affects the accuracy of process mining analyses. The lack of data synergies across systems creates further challenges, especially when synchronizing data across departments to provide a comprehensive analysis [23].

A critical aspect of process mining in healthcare is understanding, assessing, and benchmarking of key performance indicators (KPIs). KPIs directly impacting patient satisfaction include waiting time, length of stay and resource utilization. Benchmarking these factors against best practices or other departments and hospitals. This process benefits from domain expertise to validate the results, ensuring that the insights gained are practical and relevant real world context. Validation from domain experts enhances the credibility of the findings and supports the implementation of evidence based research. Expert validation is crucial in data quality and integration efforts. The involvement of healthcare professionals and administrative staff provide invaluable insights into realities of ED operations [35].

To address these challenges, this study implements the CRISP-DM (Cross Industry Standard Process for Data Mining) framework, providing a structured approach to process mining in healthcare. Key phases of the framework such as business understanding, data understanding, data preparation, modeling, and evaluation are applied using MIMIC IV ED dataset. This dataset contains de-identified health records of ED patients. The dataset undergoes through Extract-

Transform-Load (ETL) process to prepare and standardise ED data to OMOP CDM standards Figure 1. Dataset includes information on ED stay, Diagnosis, and Triage are essential to map the patient journey. Using Microsoft power automated tool Process mining techniques are applied to discover the process within the given dataset and derive necessary KPIs that are relevant for operational efficiency of ED management. Trend analysis further explores seasonal impacts, arrival process, cheif complaints and patient disposition to understand overall ED flow and bottlenecks. Domain expertise opinions are collected with real time interviews across developed, underdeveloped and developing countries.

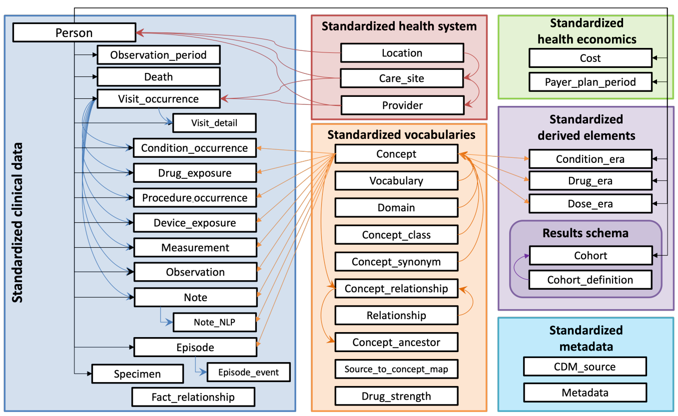

Fig. 1: OMOP CDM overview. Source: OHDSI Data Standardization.

Through this information, the study aims to improve data quality, and performance assessment in ED operations.The framework aims to also bridge the gap between data-driven insights and actionable improvements, ultimately enhancing ED performance and patient care globally.

section I provides a detail overview of the research paper section II, elaborates on research focus, detailing the inclusion and exclusion criteria, critical appraisal, and data extraction process. section III, provides an overview of the proposed research questions for this study. section IV, explains the phases of CRISP-DM methodology, introducing the business and data understanding of the MIMIC IV ED dataset. This section also covers data preparation, including the conversion to the OMOP CDM format is provided using extraction, transformation and loading (ETL) process. section V, describes the modeling phase, which identifies bottlenecks through seasonal trend analysis, arrival process, admission and stay process, and chief complaint management. Additionally, this section proposes KPIs such as case resolution rate, rework rate, patient throughput, average processing time, and seasonal impacts that is essential for validating the process mining analysis. section VI, discusses the validation of process mining analysis and insights gathered through domain expertise in developed, underdeveloped and developing countries. Finally, section VII addresses challenges, limitations encountered during data cleaning process, KPI formulation and result validation, while section VIII, summarizes the research findings and outlines future research work.

II. Literature Research

This section defines the methodological framework for deriving the research question (RQ) elaborated in the preceding sections. The methodology describes the specified criteria for database and article selection, as well as the approach to formulating research questions following the guidelines [20]. The main focus of the literature review is as follows:

1) Main focus 1: Data preparation and data quality improving techniques.
2) Main focus 2: Identification of key performance indicators (KPIs) for measuring outcomes in healthcare departments through process mining
3) Main focus 3: The role and impact of domain expertise, including the integration of expert input in the analysis.

This approach ensures that main focus points align with the study's goal of improving ED operational function by focusing on data quality, establishing relevant KPIs and leveraging the importance of domain expertise.

A. Search strategy

Using the search strategy presented by [20], the resources to formulate the research question were selected from PubMed and Scopus databases; we employed boolean operators (AND, OR) to combine keywords related to process mining, healthcare, data quality, domain expertise, key performance indicators. The keyword search and the limitations were applied as shown in Table I, which shows the process of collecting research studies for the study review.

TABLE I: Database Limitations Criteria

Key words	"Process AND mining AND healthcare AND data AND quality" OR "Process AND mining AND healthcare AND performance AND metrics" OR "Process AND mining AND healthcare AND domain AND expertise".
Source Limitations	Only Journal and conferences
Language	English
Document Type	Article, conference papers, review papers
Publication Date	>= 2015

1) Inclusion and exclusion criteria: Inclusion criteria ensure and specify the characteristics that publications must posses based on the provided keywords and concepts, the following criteria were used:

- IC1 Relevance to Healthcare: Publications must address topics related to healthcare and discuss methodologies, techniques or applications of process mining in healthcare. Additionally address concepts like KPI's, data quality and domain expertise
- IC2 Publication Date: Only publications published after 2015 are included to ensure relevance to recent advances and developments in the field.
- IC3 Case Studies: Publications presenting empirical case studies, real-world applications, or practical implementations of process mining techniques in healthcare settings are included.

- IC4 Study Types: Accepted study types include original research articles, case studies, review articles, systematic reviews, and meta-analyses.
- IC5 Language : Publications must be available in English.
- IC6 Publication Type: Both journal articles and conference proceedings are considered.

Exclusion criteria are specified characteristics that would disqualify publications if they are not of relevance;

- EC1 Lack of Direct Relevance: Publications that lack direct relevance to the research question. Studies that are unrelated to process mining or healthcare or lack applicability.
- EC2 Limited Discussion of Data Quality: Exclude papers without data quality without substantial analysis.
- EC3 Publication Types: Exclude editorials, opinions, letters, or commentaries.
- EC4 Publication Date: Publications published before 2015 to focus on recent developments in the field as the state-of-the-art literature in [30] covers the past techniques and methods.

B. Critical appraisal of collected studies

A critical appraisal was conducted to evaluate the quality of selected papers in relation to the research question outlined in section III. The quality of each paper is evaluated based on its alignment with and contribution to addressing these research questions. The quality scores can be seen in Appendix C

1) **Relevance**
   - Appraisal related to main focus point 1
     - **Data quality considerations**: Does the paper address data quality issues such as data consistency?
       Discusses data quality assessment methods and highlight the importance of standard data models?
     - Does the paper address pre and post-processing data interpretation? Highlights the data interpretation process and the methodologies employed for deriving insights from the data.
   - Appraisal related to main focus point 2
     - **KPI definition and Evaluation**: Does the paper address KPI definition and calculations to measure patient outcomes? Does it propose evaluation methods for assessing the quality of process models derived from event logs?
   - Appraisal related to main focus point 3
     - **Expert Validation**: Does the paper address expert validation derived from process mining? To what degree is the domain expertise acknowledged and utilized to enhance the interpretation and application of findings?
2) **Quality**: Assess each study's methodology as shown in Appendix B, such as data types framework, appropriate methods for data collection and analysis, KPIs metrics and domain expertise involvement.
3) **Validity**: Ensure that conclusions are well supported by data and findings are applicable.
4) **Contribution**: Assess study's impact on the field research, if it provides novel insights, develops new frameworks, or identifies trends relevant to process mining in healthcare.

For the studies, we use the rating from 0 to 4, where each quality criterion holds 1 point. The papers were qualified it is supporting the arguments that are defined, and if the minimum score was 1, the reviews were included. Appendix C ensures that study's meet the baseline quality standards and are relevant to the research questions.

C. Execution of the data extraction process and synthesis strategy

In Appendix 11, provides an overview of the study review. To begin 95 papers were picked for review from the Scopus and PubMed databases. Using Covidence, duplicated research was removed automatically, resulting in 90 papers. These papers were screened to based on inclusion and exclusion criteria to ensure each study was relevant to main focus points. The remaining 40 studies were further evaluated for eligibility, using questions such as i) Is the study related to process mining in healthcare? ii) Is the study directly addressing data assessment or quality issues relevant to research question RQ1? iii) Is the study related to one or all research questions, such as data quality, domain expertise, and/or performance? In total, 22 papers were qualified to be included in study review. The distribution of papers collected from 2015 to 2023 reflecting recent advancements in healthcare processes.

Summarizing and synthesizing the literature review findings as shown in Appendix B, gaps and key focus areas such as data quality, KPIs and the role of domain expertise are highlighted. Various studies emphasize the importance of maintaining high data quality, such as completeness, accuracy, relevance and consistency. This research sites [32] anonymizing the data correctly leads to meaningful data utilization and provides robustness to perform research. [42] addresses the challenges of extracting and syncing the data from various Electronic Health Records (EHRs) by introducing tools such as Ste package to support data extraction and event log generation.[19] addresses the use of the OMOP Common Data Model (CDM) for process mining in healthcare. CDM has been proven to be valuable in the healthcare sector; however, it needs to be supplemented with data sources for outpatient and emergency room process. [12] uses Adoption Readiness Assessment and Maturity Model (RAMM) to improve data management in healthcare sectors. This framework helps in assessing organization's data management maturity and recommends improvements. In terms of key performance metrics [21] suggests using QUAD metrics (Fitness, Precision, Simplicity and Generalization) as a universal standard for evaluation process models in healthcare. In case studies presented in [2] from Egyptian healthcare facilities, demonstrating the impact of implementing bottleneck detection to improve outpatient clinic management used the analysis of event logs and the need for domain expertise.

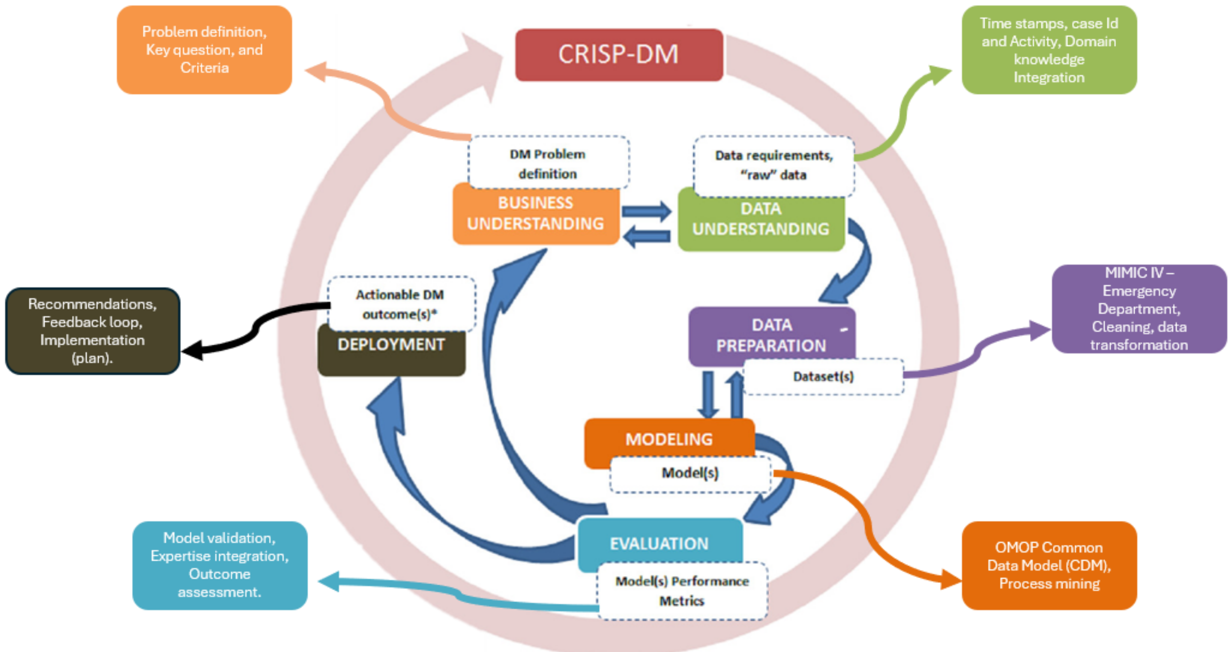

Fig. 2: CRISP DM Framework. Source: ScienceDirect.

[7] also identifies the process improvement opportunities, and stresses the inputs of domain expertise a crucial step in interpreting and applying process mining for effective results. Finally, studies like [18] outlines difficulties such as limited access to patient data due to confidentiality, data quality issues like MIMIC-III and the complexity of analyzing patient care processes. These challenges highlight the need for data preprocessing to avoid incomplete analyses.

III. Research Questions

The following research questions were formulated based on the gaps identified during the literature review.

1) What frameworks and methodologies can ensure data accuracy, completeness and consistency, to enhance process mining analyses?
2) Which KPIs can be derived to identify bottlenecks, evaluate operational efficiency and patient outcome using process mining techniques?
3) Can domain expertise across globe validate process mining insights through real world experiences?

IV. Methodologies

In this section, framework for organizing and structuring the research are outlined. [39], explains CRISP-DM (Cross Industry Standard Process for Data Mining) has shown in Figure 2 is implemented.

CRISP-DM is an iterative, structured, and flexible framework that makes it an ideal choice for conducting process mining in healthcare, especially in emergency care. It addresses challenges of healthcare data, aligns technical and clinical objectives that allow for iterative improvements of models, ensuring findings can be applied effectively in real-world settings. Using CRISP-DM the research ensures a thorough, repeatable and scalable approach to improving the healthcare process. The six phases are described in the following steps:

1) Business Understanding : This in an initial phase that summaries the research question as shown in section II and provides motivation of the research objectives from a business perspective in subsection IV-A.
2) Data Understanding: Second phase, relevant data is collected to gain insights and understanding the data characteristics, quality, missing values, or pattern that is useful for the research analysis, as mentioned in subsection IV-B that answers research question 1.
3) Data Preparation: The third phase involves data cleaning, transformation and one the most time consuming as it is the implementation phase of the OMOP CDM format to answer research question 1.
4) Modeling: The fourth phase involves selecting suitable modeling techniques for the prepared data, as mentioned in section V. This phase also includes finding tuning model parameters, evaluating the best model and approaches providing solution for research question 2.
5) Evaluation: The fifth phase focuses on assessing the model against business objectives defined in the as mentioned in section III. This phase also involves ensuring the required standards that will provide meaningful insights when deployed in section VI and answers research question 3.

TABLE II: Summary Of Key Aspects Of The MIMIC-IV Emergency Department Dataset

Aspect	**Details**
Data Source	Beth Israel Deaconess Medical Center's Emergency Department (ED)
Number of Patients	Approximately 80,000 unique ED visits (varies with data updates)
Number of Visits	Over 100,000 ED visits
Time Span	Data from 2008 to 2019 (for MIMIC-IV)
Data Types	- Patient Demographics: Age, sex, race, etc. - Visit Information: Triage notes, admission times - Clinical Data: Vital signs, laboratory results - Diagnoses: ICD-9/ICD-10 codes - Procedures: Documented procedures and interventions
Diagnosis Codes	ICD-9 and ICD-10 codes
Vital Signs	Heart rate, blood pressure, temperature, respiratory rate, oxygen saturation, etc.
Laboratory Results	Blood tests (e.g., CBC, BMP), urine tests, and other relevant diagnostics
Clinical Notes	Free-text notes from triage, physician assessments, and discharge summaries
Admission Details	Time and date of admission, discharge details, and transfer information
Common Conditions	- Sepsis - Chest pain - Respiratory distress - Abdominal pain - Head trauma
Procedures Recorded	Common ED procedures (e.g., intubation, central line placement, wound suturing)
Data Frequency	Varies: vital signs recorded at intervals (e.g., every hour), lab results based on tests ordered
Access and Use	Publicly available with data use agreements; requires adherence to ethical guidelines and IRB approvals

A. Business Understanding

By leveraging domain expertise and addressing real challenges in the healthcare sector, as identified through the literature review in section II, this research aims to tackle the dynamic nature of Emergency Department (ED) operations, which are influenced by factors various factors. The primary objective is to enhance data understanding and quality by mapping to OMOP CDM vocabularies as shown in Figure 1.

The second research question evolves on bottleneck identification, patient flow efficiency, and standardise KPIs to assess process mining on healthcare outcomes [29, 1]. The final question emphasizes on the role of domain expertise in validating data mapping outcomes and process mining results, bridging gaps between data-driven insights and practical improvements in emergency departments [24, 6].

B. Data Understanding and Preparation

This section addresses first research question detailing the type of dataset selected, and providing an overview. To enhance data quality and consistency, framework such as OMOP CDM has been applied. The implementation leverages Extract, Transform and Load process to structure the data and align with CDM standards.

*1) **Data Understanding**:* One of the main focus of this research is to leverage process mining implementation to identify bottlenecks and optimize patient outcomes in the ED, making data selection and understanding essential.

The [17] MIMIC-IV ED dataset is a collection of de-identified health records from the emergency department of Beth Israel Deaconess Medical Center. The dataset includes information such as patient demographics, vital signs, laboratory test results, procedures, medications and patient outcomes as shown in Table II. This records provide insights into ED operations, patient flow, and diagnosis, making the dataset highly valuable for studies in process optimization and healthcare quality improvement.

MIMIC-IV ED consists of *edstays, diagnosis, triage information,has the vital signs, medication taken prior to ED stay and medication dispensed during ED stay* as shown in Figure 3. The goal of this research is to focus on patient journey using *edstays, diagnosis and triage* which has information of start and end times, by determining severity of the diagnosis. Understanding the key variables in the dataset can provide critical insights into the nature of data and its characteristics, which is fundamental for analysis.

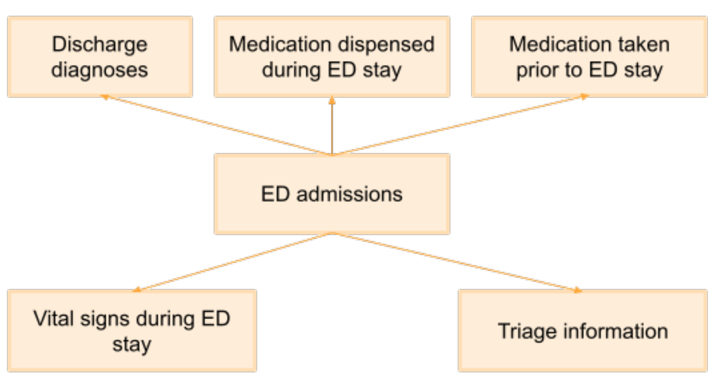

Fig. 3: MIMIC IV ED Data Structure.

- ED STAYS Dataset: The ED STAYS overview provides key metrics such as number of records, unique patients, average and median length of stays, and the most common patient disposition as shown in Table III. Understanding the distribution of length of stay as shown in Figure 4, arrival patterns, and patient demographics is critical for assessing patient flow and resource utilization in the ED. The average length of stay (LOS) according to the figure is 7 hours, which explains that many cases are dismissed, sent home or transfered to another department for further diagnosis.

TABLE III: EDSTAYS Overview

Metric	EDSTAYS Overview
Number of Records	425,087
Number of Unique Patients	205,504
Average Length of Stay (hours)	7.16
Median Length of Stay (hours)	5.47
Most Common Disposition	HOME

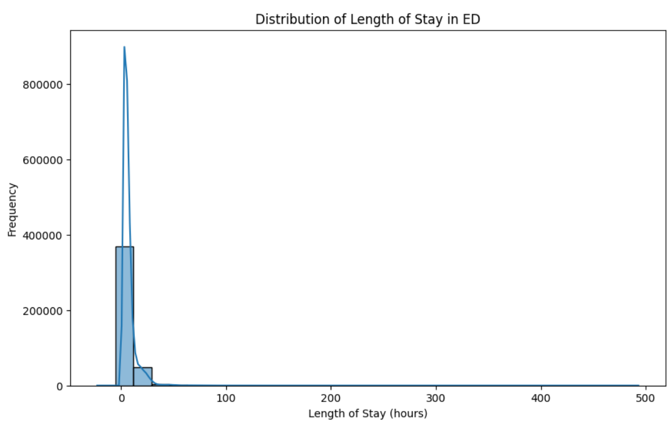

Fig. 4: ED STAYS Length Of Stay Distribution.

- DIAGNOSIS overview summarizes International Classification of Diseases (ICD) codes and description assigned to patients during the ED visit as shown in Table IV. This data helps in understanding the prevalence of various conditions treated in the ED; identifying the common and critical health issues enables, as shown in Figure 6, with the top 10 diagnosis correlate with patient outcomes.

TABLE IV: DIAGNOSIS Overview

Metric	DIAGNOSIS Overview
Number of Records	899,050
Number of Unique Diagnoses	13,178
Top Diagnosis	HYPERTENSION
Most Common Diagnosis Count	26,816

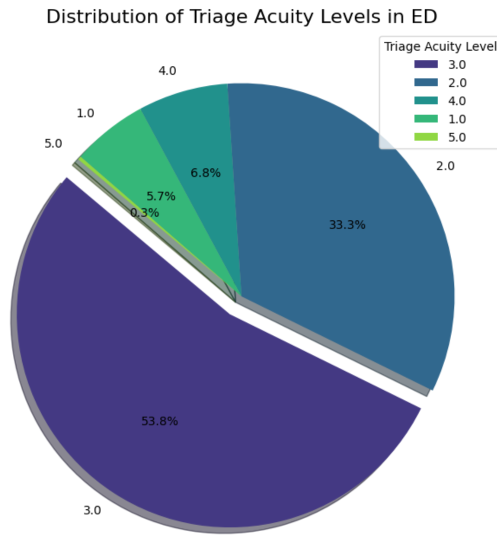

Fig. 5: Triage Acuity Level.

- TRIAGE is the preliminary assessment of patients to determine the urgency of their needs for treatment. The overview showcases the initial assessment made with the scale of 1-5. when patients arrive in the ED, such as the most common triage acuity level as shown in Figure 5, is 3.0, a large percentage of medium severity of issues that arrives. The second largest severity is 4.0 and above which can lead to transfer or admitting in the hospital. This helps in understanding severity of their conditions, and to plan the patient journey using Table V.

TABLE V: TRIAGE Overview

Metric	TRIAGE Overview
Number of Records	425,087
Number of Unique Patients	205,504
Most Common Triage Acuity Level	3.0
Average Heart Rate	85.08
Median Systolic Blood Pressure	133.0
Median Respiratory Rate	18.0
Median Temperature	98.0

2) Data Preparation: In CRISP-DM methodology, data preparation phase involves transforming data to be ready for modeling. Data is mapped to the Observational Medical Outcomes Partnerships (OMOP) Common Data Model (CDM) vocabularies, to ensure data consistency, interoperability, and standardization [34]. The OMOP CDM is a standardized data model designed to provide analysis and share observational outcomes of healthcare data across various domains, such as patient demographics, observations, drug exposure, procedures, and diagnoses. The OMOP CDM with Figure 1, consists predefined schema that act as a representation of healthcare data and vocabularies that is used to standardize terms across different datasets.

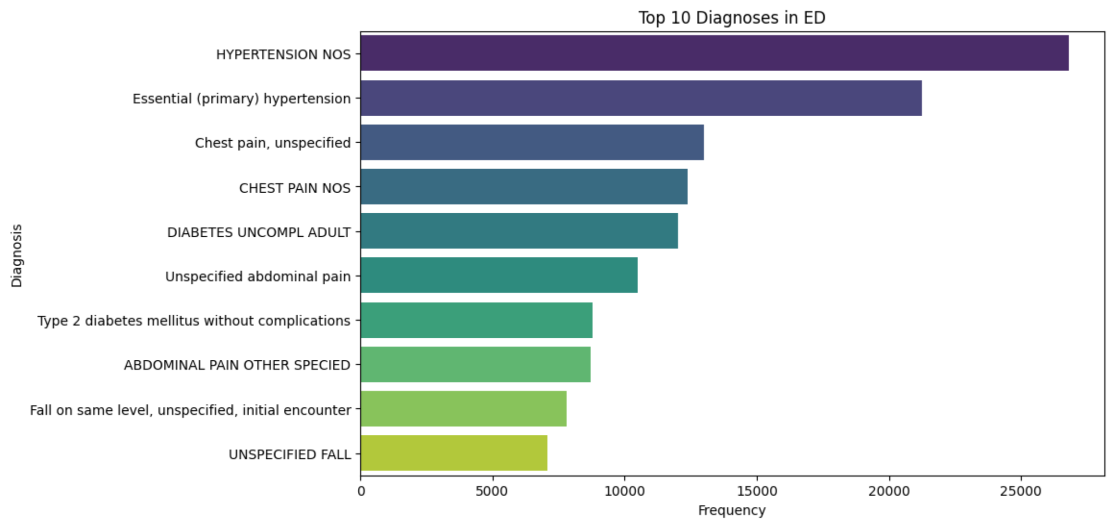

Fig. 6: Top 10 Diagnosis.

TABLE VI: Mapping of MIMIC-IV Emergency Department Data to OMOP CDM Format

MIMIC-IV Table	MIMIC-IV Column	OMOP CDM Table	OMOP CDM Field	Mapped Concept
edstays	subject_id	visit_occurrence	person_id	Person Identifier
edstays	hadm_id	visit_occurrence	visit_occurrence_id	Visit Identifier
edstays	intime	visit_occurrence	start_datetime	Visit Start Time
edstays	outtime	visit_occurrence	end_datetime	Visit End Time
edstays	admittime	visit_occurrence	admitting_source_concept_id	Admitting Source Concept
edstays	dischtime	visit_occurrence	discharge_to_concept_id	Discharge Disposition Concept
diagnosis	subject_id	condition_occurrence	person_id	Person Identifier
diagnosis	hadm_id	condition_occurrence	visit_occurrence_id	Visit Identifier
diagnosis	icd_code	condition_occurrence	condition_source_value	Condition Source Value (ICD Code)
diagnosis	icd_code (mapped)	condition_occurrence	condition_concept_id	Condition Concept ID (Standard Concept)
diagnosis	charttime	condition_occurrence	condition_start_datetime	Condition Start Time
diagnosis	seq_num	condition_occurrence	condition_type_concept_id	Condition Type Concept (Priority)
triage	subject_id	observation	person_id	Person Identifier
triage	hadm_id	observation	visit_occurrence_id	Visit Identifier
triage	level	observation	observation_concept_id	Triage Level Concept
triage	charttime	observation	observation_datetime	Observation Time (Triage Time)
triage	hr, bp, etc.	measurement	measurement_concept_id	Measurement Concept (e.g., Heart Rate)
triage	value	measurement	value_as_number	Measurement Value
triage	unit	measurement	unit_concept_id	Measurement Unit Concept

The aim of mapping MIMIC-IV ED datasets (EDSTAYS, DIAGNOSIS and TRIAGE) to the appropriate tables and fields is to standardize data for analysis, the mappings are shown in Table VI, and Figure 7. The OMOP CDM mapping process is performed by executing custom Python scripts using Extraction-Transformation and Loading (ETL):

1) **Extraction** : Relevant data was extracted from the MIMIC-IV ED tables using Python script. These scripts are designed to extract fields necessary for mapping to OMOP CDM concepts, i.e., diagnosis codes, visit identifiers, patient demographics, timestamps, vital signs and initial assessment information.
2) **Transformation**: All timestamps were converted to OMOP CDM's datetime format using Python's datetime library.
   - ED Stays: ED visit data mapped to *visit occurrence* table in the OMOP CDM, key elements are visit start and end time, visit type, patient id and movements within ED.
   - Diagnosis: ICD codes in diagnosis are mapped in *condition occurrence* , OHDSI Athena Vocabulary System [33] enabling diagnosis codes mapping from MIMIC IV to the OMOP CDM standard vocabularies such as (ICD 9, ICD 10, SNOMED).

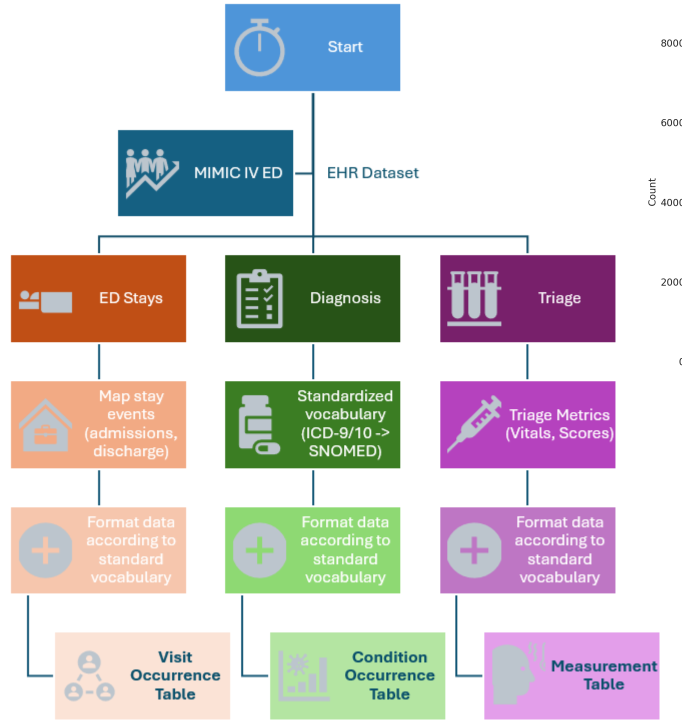

Fig. 7: Data Preparation And Mapping

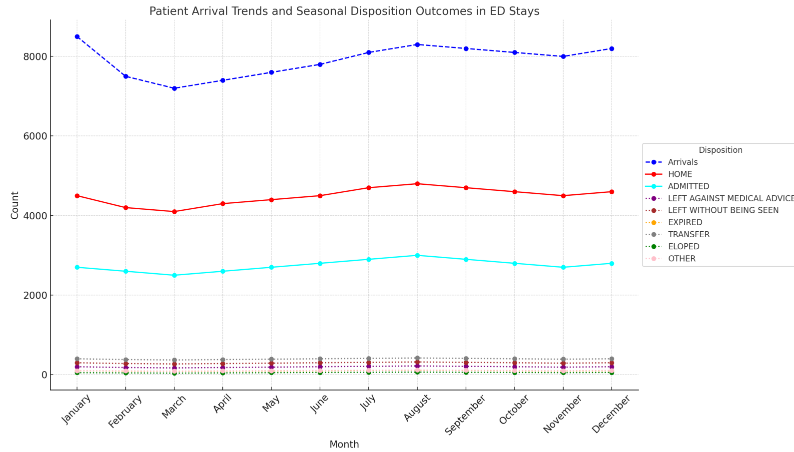

Fig. 8: Seasonal Trends Using Arrival And Disposition.

   - Triage: Measurements of triage assessments were mapped to **observation** entries, chief complaints to *measurements and observation* table.

3) **Loading**: Transformed data was loaded into corresponding the OMOP tables (condition occurrence,visit occurrence, measurement, observation) to ensure consistency across tables, and standard representation. Loading process includes mappping between tables to preserve integrity for patient ID, diagnoses and timestamps.

4) **Data quality checks** : Python scripts were used to ensure no critical data fields were missing after the ETL process, particularly for key identifiers and timestamps. Data was also validated for consistency using Python ensuring all the diagnosis codes are mapped to OMOP concept IDs.

V. Process Modeling

In this section, we identify and analyze the bottlenecks within the healthcare processes using Power Automate tool by Microsoft. Process mining implementation helps identifying and, proposes standard KPIs to monitor operations and ensure the best possible care in a timely manner.

- **Seasonal Trend Analysis**: It is a method used to identify fluctuations in data that occur at regular intervals with months, quarters or years. The trends are driven by external, time stamps such as weather and holiday periods. As shown in Figure 8, the seasonal trends in patient arrival and disposition outcomes in ED show significant fluctuations throughout the year. Patient arrivals peak in winter, especially in the month of January, with a dip that is followed by early spring (February and March). Volumes stabilize over the summer with a slight rise in December, due to possible seasonal illnesses like flu. Disposition outcomes such as admissions, discharges, and transfers, reflect these trends, although admission rates remain consistent.
- **Arrival process** : The process to identify the mode of transport opted by patients. As shown in Appendix 12, 'Ambulance' and 'Walk In' nodes have high rework and self-loop counts that show inefficiencies in the initial patient intake process. The 'Unknown' node also exhibits significant looping, suggesting unclear pathways that may lead to delays in patient management. These patterns highlight the requirement for streamlined processes to enhance patient flow and reduce bottlenecks in the intake phase.
- **Admission and Stay (Disposition) process** : The journey of patients tracked post triage phase. Appendix 13, shows that the 'Admitted', node, acts as a central part of patient flow within the facility, indicating a high volume of cases with frequent rework. This points to bottlenecks resources that is leading to delays in patient care progression. Nodes related to 'Transferred' and 'Eloped' show steady inflows and outflows, raising concerns about potential inefficiencies in patient management.
- **Cheif Complaint Management** : Nodes with common complaints such as 'ETOH' (alcohol abuse),'s/p Fall' and 'Abd Pain' exhibit high rework and self-loop rates as shown in Appendix 14. The 'ETOH' node demonstrates a rework rate of 94% with 68% of cases involving self-loops, suggesting that these cases are frequently revisited without resolution. Nodes with 'SI' (Suicidal intentions) and 'Altered Mental Status' where the complexity of these cases leads to multiple interactions and extended patient stays, indicating bottlenecks in these conditions.

To provide solution for research question 2, KPIs are derived by combining lean management, process mining methodologies and domain specific practices in healthcare performance evaluation . The chosen KPIs measure critical aspects of patient care that directly impact both efficiency and effective care in time as shown in Table VII.

- **Rework Rate**: measures the frequency of cases requiring reprocessing, indicating potential bottlenecks. Current rework rate is between 75% to 94%. Below 10% is recommend according to lean management principles which emphasize minimal waste, enhance efficiencies and improving process flow [44, 25]
- **Average Processing Time**: Reducing patient time spent in each stage is critical for improving throughput and patient outcomes. [41, 15] The target of a 20% to 30% reduction is suggested according to lean principles to achieve optimal processing times.
- **Patient Throughput**: This is to measure how many patients a healthcare facility can effectively process. Maintaining or increasing throughput is essential, especially in emergency departments. [26, 14] Optimizing throughput is vital for balancing demand and capacity in healthcare which can lead to better resource utilization.
- **Case Resolution Rate**: [4, 10] If 90% of the cases are handled correctly the first time, reducing delays and improving patient satisfaction can be indicators of an effective process in healthcare unit. The case resolution reflects the percentage of cases needing rework or looping back. In this research, often below 70% are in complex cases.

VI. Result Evaluation

The key findings from developed and developing/non-developed countries reveal significant differences in ED management, that are related to the specific key performance indicators as a part of the result validated by the domain expertise from countries such as Spain, India, Argentina, Ecuador and South Africa as shown in Figure 9 and Table VIII. The validated results also confirm that our findings using MIMIC IV ED dataset is closely align with the real time functioning of the healthcare landscape. This proves that our studies can be implemented under the guidance of domain expertise.

- **Rework rate and Average processing time** : As mentioned in previous sections of dataset overview, MIMIC IV ED sheds light on the structured, standardized processes, especially for common conditions like sepsis that leads to lower rework rates and improved efficiency in patient management.Developed countries have structured workflow and reduced rework rates when compared to developing/non developed countries that lack structure and are less standardized, facing higher rework rates.
- **Patient Throughput** : The dataset demonstrates high patient throughput is associated with optimizing staffing and resource management, especially during peak hours. Use of triage and resource allocations can reduce wait times. Developed countries align with findings of the dataset by maintaining higher throughput and resource management using triage systems such as the Manchester and Australian triage systems. However, developing/non developed countries are impacted by overcrowding and limited staffing without the use of protocols or triage usage.
- **Case Resolution Rate** : MIMIC IV ED and developed countries shows higher case resolution rates with well document, repeatable workflows and use the digital records such as ICD codes for diagnoses and treatments. developing/non developed countries showcase less structured and reliance on digital records, and face challenges in meeting higher resolution rates.
- **Seasonal impact on ED performance** : Seasonal data trends in the dataset show that patient flow and case types fluctuate with seasons as winter approaches higher respiratory cases are checked in and in summer, higher rates of strokes due to local tourism. Developing countries face significant pressure on EDs during the seasonal fluctuations with complaints of fever, respiratory illness, esophageal gastric/heart attack, infection; the lack of strategies and workflow lead to longer processing time and patient flow.
- **Patient Journey**: Patients in developed countries generally benefit from more efficient triage, fast diagnosis such as CT scans, MRI and lab tests, with quick results. Defined treatment options, and streamlined processes for discharge with disposition process (discharge, admission, or transfer) often supported by electronic systems. In contrast, developing/non-developed countries usually encounter limited access to diagnosis which in many case might need transfer to other departments or facilities for testing, results can take longer. Treatments are delayed due to constraints, such as limited medication stocks, patients may need to undergo long waiting times with big crowds and variability in critical care resources. This cause delay in the decision making process of disposition and leading to chaos in assigning proper care.

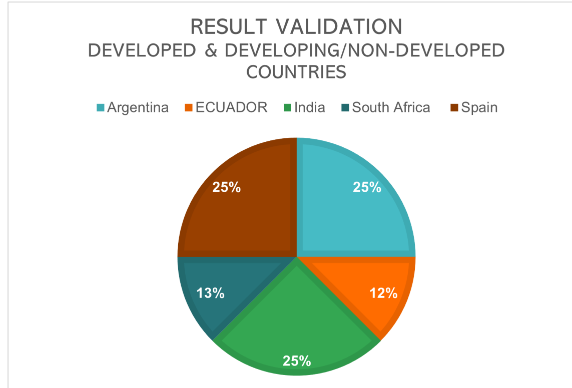

Fig. 9: Result Validation Responses

TABLE VII: Key Performance Indicators (KPIs) with Current Outcomes and Proposed Targets

KPI	Description	Current Outcome	Proposed Target
Rework Rate	Measures the frequency of cases that require rework at each process stage. A higher rate indicates inefficiencies.	Ranges from 75% to 94% across nodes	Below 10%
Average Processing Time	Monitors the time patients spend in each process stage, with a focus on reducing delays.	Based on the LOS data, the average is 7 hours	Reduction by 20%
Patient Throughput	Counts the number of patients processed by each node, ensuring that resources match demand.	Varies by node; critical nodes often exceed capacity	Maintain or Increase
Case Resolution Rate	Evaluates the percentage of cases resolved without needing rework or looping, indicating process efficiency.	Varies significantly; often below 70% in complex cases	Above 90%

TABLE VIII: Comparison of KPI Performance between Developed and Developing Countries

KPI's and Trends	Developed Countries	Developing Countries	KPI Correlation
Rework Rate & Average Processing Time	Structured workflows, standardized protocols, lower rework rates, shorter processing times.	Higher rework rates due to lack of structured workflows, longer processing times.	Developed countries likely meet rework rate ¡10% and reduced processing times; Developing countries need workflow standardization and EHR integration for improvement.
Patient Throughput	Efficient throughput due to better staffing ratios and resource utilization.	Throughput affected by overcrowding, inadequate staffing, and inconsistent arrival methods.	Developed countries meet high throughput KPIs; Developing countries need strategies to balance demand and optimize patient flow.
Case Resolution Rate	High case resolution rate due to structured workflows and fewer interruptions.	Lower resolution rates due to long wait times, poor workflow, and lack of digital tools.	Developed countries approach 90% resolution rate; Developing countries need workflow and digital system improvements.
Seasonal Impact on ED Performance	Seasonal trends managed through standardized processes, minimizing delays.	Seasonal fluctuations create pressure due to lack of preemptive strategies and workflows.	Developed countries manage processing time efficiently; Developing countries need better forecasting and standard workflows.
Patient Journey	Smooth, rapid diagnosis and treatment due to triage and digital data collection.	Resource constraints due to lack of structure and planning in healthcare processes.	-NA-

VII. Discussion and Limitations

During the data cleaning process, several challenges with data quality and cleaning occurred. Inconsistent data entry, missing data, data anomalies and outliers, and standardization of ICD codes to OMOP CDM format to define KPIs were challenging. Data quality issues were seen in developed and developing/non-developed countries but are more pronounced in settings with less standardized data collection processes. Results clearly state there is significant variation in ED performance and process efficiencies between developed and developing/non developed countries. Developed countries showcase structured workflows, better resource allocation and use of digital tools that contributes patient care, lower rework rates and improved case resolution rates that align with key performance indicators (KPIs). Whereas results of developing countries present challenges like overcrowding, insufficient staffing, and fragmented workflows leading to higher rework rates and lower case resolution rates. The lack of digital tools and standardized protocols further complicate data harmonization contributing to delays in achieving accurate KPIs.

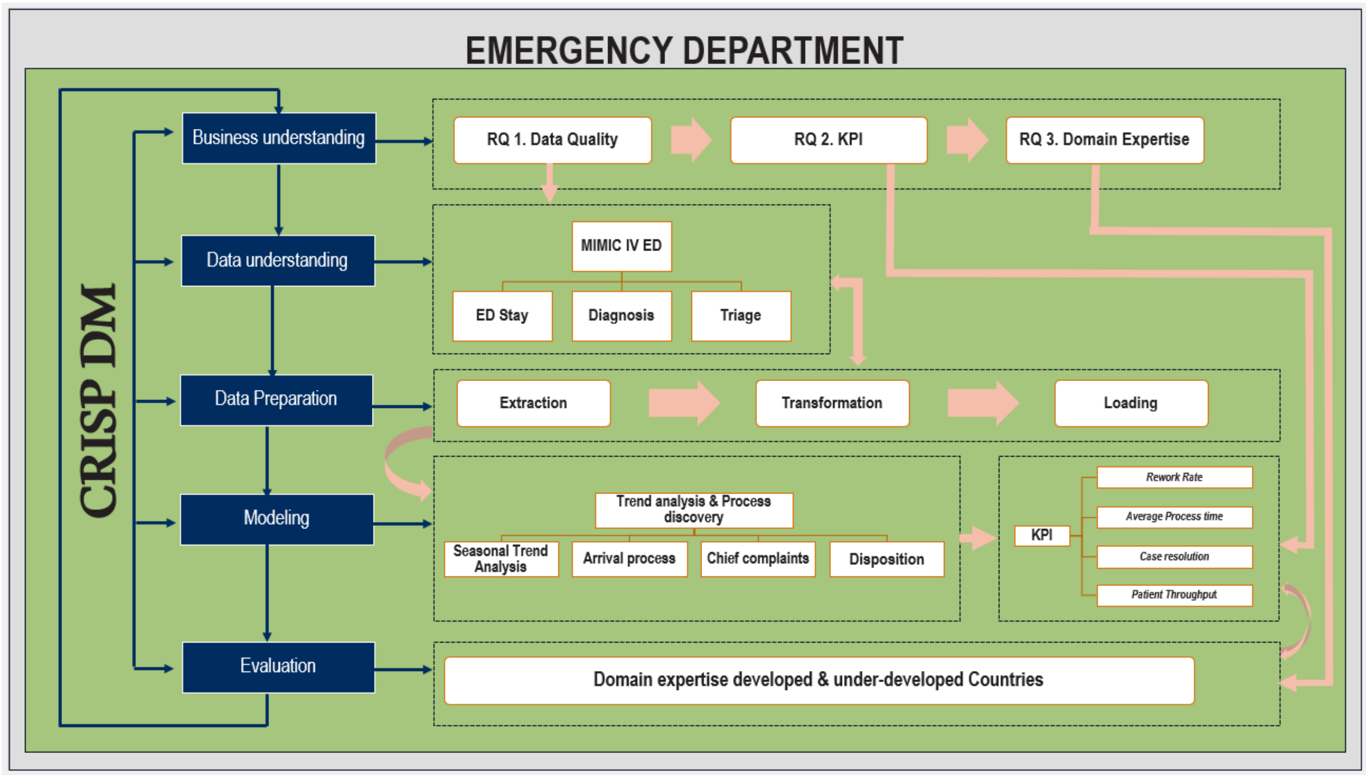

Fig. 10: Research Overview.

Temporal variability such as seasonal trends also impact ED performance with all the types of countries that are noted to be fluctuating in patient arrival. However, developed countries have an efficient seasonal management and protocols to handle the variability. Enhancing data standardization can be crucial for developing countries to support data consistency and interoperability across systems. Implementation of frameworks like OMOP CDM can help with standardization's across globe by providing uniform medical codes and protocols to manage healthcare organizations. Involve domain expertise in validation for enhancing workflows by using there expertise in validating data quality, and practical feasible clinical settings.

VIII. Conclusion and Future work

As shown in Figure 10, the study highlights value of process mining in healthcare to improve overall performance of an ED. The research applies the **CRISP-DM framework** with six phases - **Business understanding, Data Understanding, Data Preparation, Modeling and Evaluation** to ensure a structured approach. The study focuses on three key research questions: **Data quality, KPIs and the role of domain exepertise**. MIMIC IV ED dataset was mapped to **OMOP CDM** framework to ensure data quality, accuracy and completeness. This was performed using the process of data extraction, transformation, and loading (ETL). Process mining techniques using power automate by Microsoft was used for **process discovery** to analyse critical workflows like arrival process, seasonal trends, chief complaints, and patient disposition. Using the process analysis and lean management principles KPIs such as **rework rate, process time, case resolution and patient through** were derived to align with the healthcare workflows. Result validation was performed by the domain expertise in countries like **Spain, India, Argentina, Ecuador and South Africa**. This offered valuable insights into the challenges faced in real world scenarios in the ED of developed, and developing/non developed countries. The study makes valuable contributions by demonstrating how clean data and domain expertise can enhance ED management. Additionally, highlighting the importance of frameworks like OMOP CDM to improve data standardisation and interoperability across global healthcare systems.

Healthcare organizations can benefit from implementing advanced process mining techniques such as conformance checking and should be established as best practices. Studies can be further explored to analyze trends in Emergency cases due to race, gender and landscape/geographic regions, as these factors can influence overall emergency care setup. Domain expertise have their specialists and skills based on the population they serve, as these are shaped by demographics and environmental factors. With increasing globalization and movement of immigrants across continents, healthcare systems need to evolve the disease patterns, and consider genetic factors that may arise in diverse populations. By addressing these regional factors, healthcare organizations can be better prepared for the varying needs of their population and enhance the emergency care experience by ensuring their staff is equipped with expertise required to manage complex and evolving healthcare needs.

Appendix A
Literature & Process Discovery Visualizations

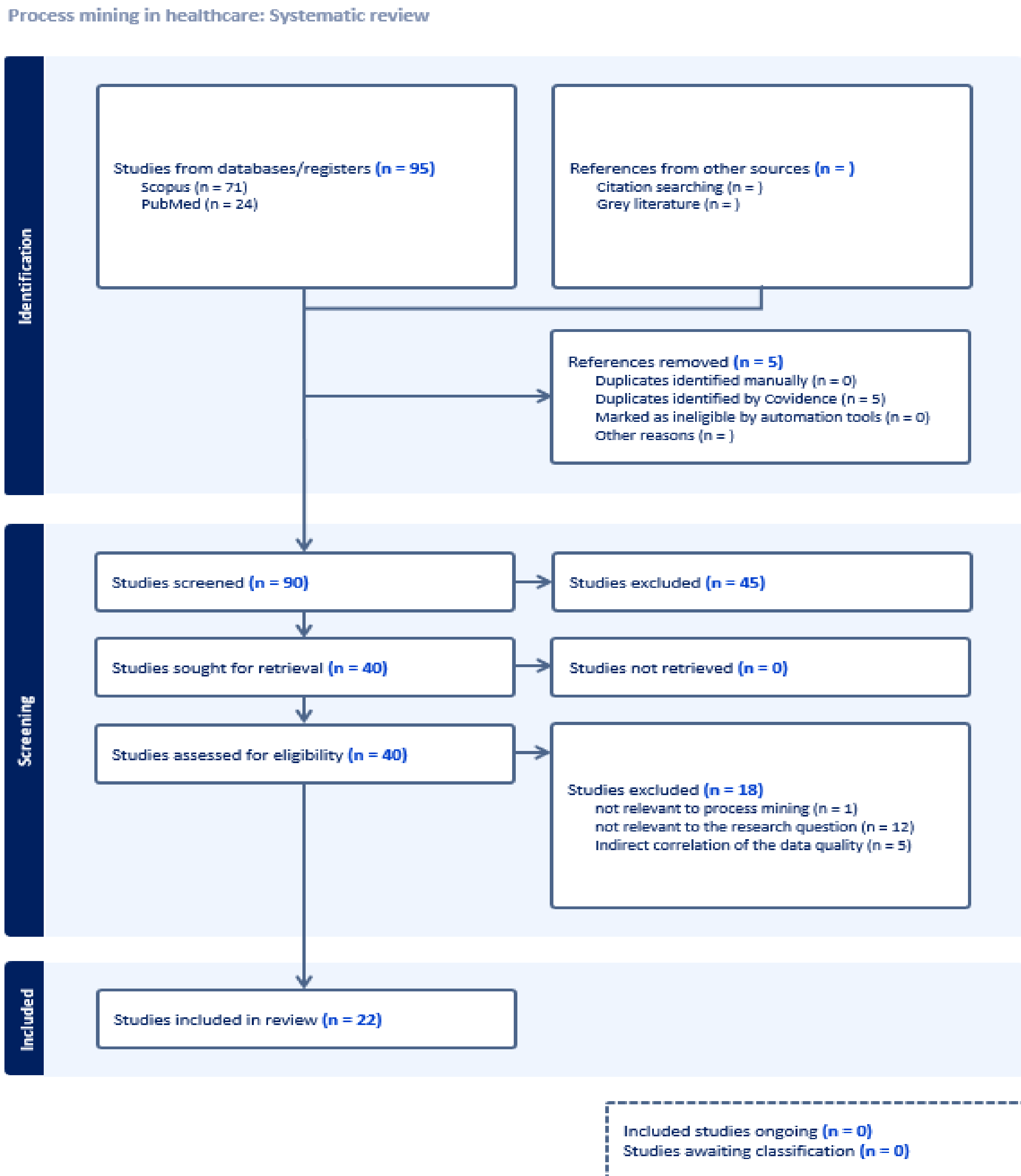

Fig. 11: PRISMA Flow Diagram.

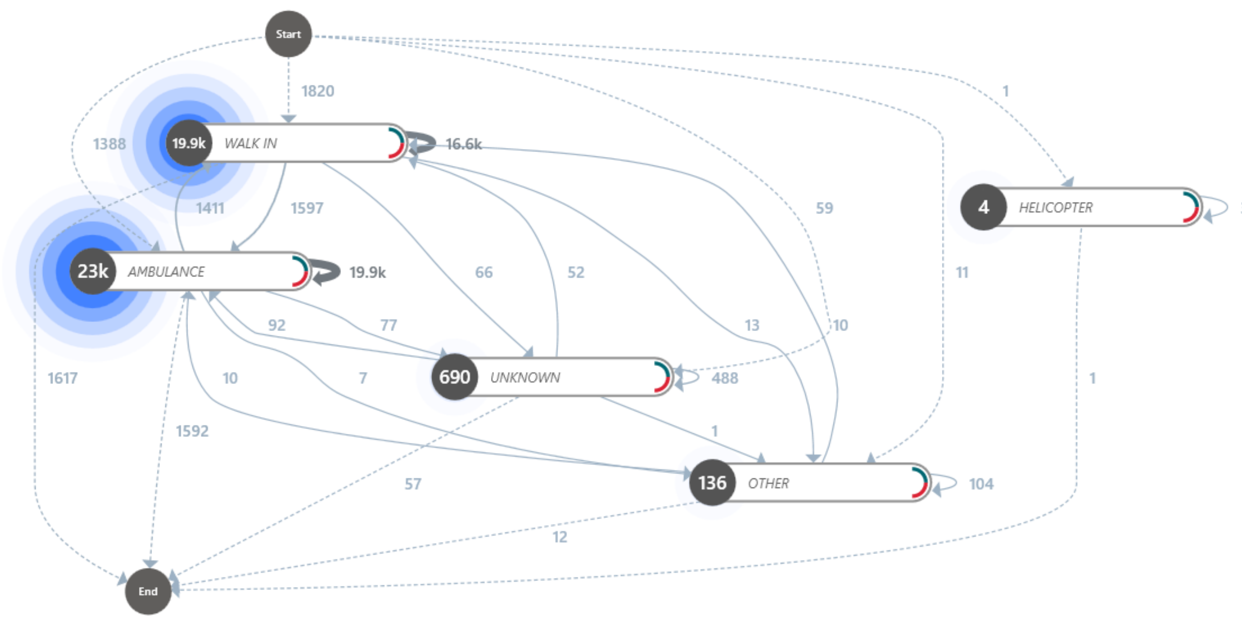

Fig. 12: Arrival Process Discovery.

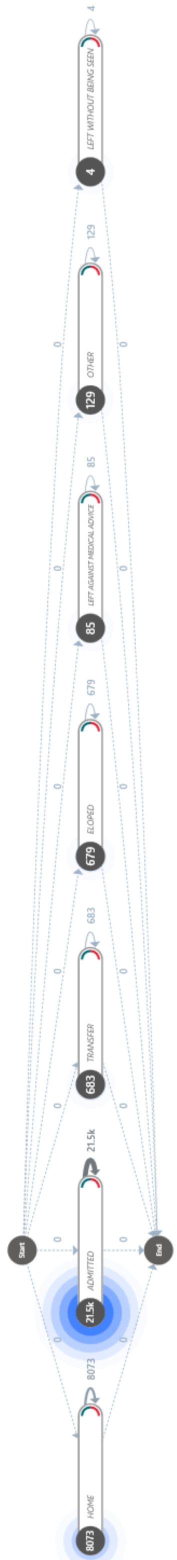

Fig. 13: Disposition Process Discovery.

Fig. 14: Chief Complaints Process Discovery.

Appendix B
Study Overview

Research Question	Tool	Methodology	Experiment settings	References
RQ1	ProM	L* lifecycle model for data quality assessment	Reconstruction of the MIMIC-III dataset into a relational database in PostgresSQL, which included downloading 26 csv files and importing the data into the PostgreSQL database.	[32]
RQ1	Common Data Model (CDM)	process discovery, conformance checking, performance analysis, and patient journey analysis	Dataset collect from various sources, Korean tertiary hospital, including Total Laparoscopic Hysterectomy (TLH), Total Hip Replacement (THR), Coronary Bypass (CB), Transcatheter Aortic Valve Implantation (TAVI), and Pancreaticoduodenectomy (PD)	[19]
RQ1	Data Management Maturity Model (D3M& Care Pathway Data Quality Framework (CP-DQF) for formal assessments	Care Pathway Data Quality Framework (CP-DQF) developed by Fox et al. EDQ improvement. Readiness Assessment and Maturity Model (RAMM) to develop organizational change management	Assessing event data quality (EDQ) in healthcare organizations and care pathway data quality	[12]
RQ1 & RQ3	None	Data-Driven Process Simulation (DDPS) for healthcare capacity management	The study utilized two years of process execution data, from March 2017 to March 2019, extracted from the hospital's Radiology Information System (RIS) to build the simulation model. Additionally, a validation dataset from March 2019 to March 2020 was used to validate the model.	[16]
RQ1	None	Fluxicon's Disco process mining tool	The dataset comprised anonymized data from 708 patient stays over a one-year period, with a total of 185,496 single data items	[43]
RQ2 , RQ3	None	Process Performance Measurement, Framework for Process Performance Indicators (PPIs), Performance Analysis	The study utilized real-life clinical event logs collected from a tertiary hospital in Korea. The event log included approximately 460,000 events for about 30,000 patients who visited the emergency room.	[5]
RQ1	ProMLite 1.2	Process Mining Project Methodology (PMPM), CRISP-DM	Queensland Ambulance Service (QAS) and Retrieval Services Queensland (RSQ) datasets.	[37]
RQ2	None	MEG (Mining of Events using Genetic) algorithmevolutionary algorithms, genetic algorithms, Petri nets, causal matrix, workflow nets, tabu search, Monte Carlo optimization, and MXML legacy classifier	Extraction of event logs from legacy information systems	[38]
RQ2	PM4PY and ProFIT toolkits	creation and implementation of the QUAD metrics	Acquisition of datasets related to the Central Venous Catheter Process and Sepsis Treatment Careflow	[21]
RQ1 & RQ3	ProM, Inductive visual Miner, and the ProcessProfiler3D	The Inductive visual Miner was selected for its suitability for discovery, conformance, and comparative performance analyses, as well as its excellent data filtering capabilities and robustness in dealing with complex models	The Queensland Ambulance Service (QAS), Retrieval Services Queensland (RSQ), Emergency Department Collection (EDC), Queensland Hospital Admitted Patient Collection (QHAPDC), and Births, Deaths, and Marriages Data (BDM). The data was linked by Qld Health's Statistical Services Branch Data Linkage Unit, resulting in a patient identifier being added to each record of each individual source data set	[7]
RQ1 and RQ2	pMineR	use of PWL (Pseudo Workflow Language) for conformance checking, leading to the development of CSL as a more flexible and principled approach for data gathering and event log generation.	Real-world data from 12,000 patients of the Gemelli Hospital	[42]
RQ1, RQ2 & RQ3	ProM	algorithms include the alpha-algorithm, heuristic miner, inductive miner, ILP miner, and ETM miner	Patient careflows in the cardiac surgery unit of an Egyptian hospital.	[27]
RQ1, RQ2 & RQ3	None	Process Mining Project Methodology in Healthcare (PM2HC) in the context of healthcare processes	Unit of the Clinical Analysis Laboratory at the Hospital de Clínicas (ULAC) data collection from the laboratory	[28]
RQ2	Celonis Execution Management System (EMS) platform for KPI visualization and analysis	The methods discussed in the document involve the integration of customer journey mapping (CJM) with process mining to optimize the creation of healthcare indicators.	The data for the experiment is extracted from a Hospital Information System, specifically from the Emergency Department's Electronic Medical Record from a teaching hospital in Santiago, Chile.	[11]
RQ2	Healthcare 4.0 systems	use of statistical process control (SPC) charts and the establishment of theoretical fundamentals required for survival analytics to evaluate the performance of unlabelled datasets in real-time.	The experiment analyses process data from 560 patients, generating 1450 alerts. The study uses a time window called the Analysed Post Alert Time-Window (APA), which is sliced into 24 consecutive blocks of 15-minute time intervals to determine the genuineness of the generated alerts.	[31]
RQ1	Arena-Rockwell Automation version 16.20.00	data-driven business process simulation and bottleneck detection in healthcare.	The experiment setup in the document involves the use of two years of data from the Clinic Information Systems (CIS) in Egypt, spanning from January 2019 to December 2020, to build a discrete event simulation (DES) model	[2]
RQ1	ProM, Disco, and PALIA	Fuzzy Miner, Inductive Visual Miner, and Trace Clustering,	Literature review of clinically-relevant case studies in healthcare process mining from 2016 to 2018	[22]
RQ1	ProM 6.5.1	L* lifecycle model, Performance/Conformance Analysis, Process Mining Project Methodology (PM2) for data processing and extraction	The MIMIC-III dataset for process mining in oncology	[18]
RQ1	ProM, DECLARE, DISCO	L* life-cycle model, race Clustering, Sequence Analyzer, Fuzzy Miner, Alpha Miner, Genetic Miner, Heuristic Miner, and Conformance Checker	Geographical analysis and classification of case studies have been performed, with a concentration of studies in Europe and a limited number in North America, Asia, and Australia. The case studies have been divided and classified into various medical domains, such as cardiology, caregiving processes, dentistry, diabetes, intensive care unit, medication, oncology, radiotherapy, and surgery.	[40]
RQ2	ProM, and jPM software	Process remodelling or analyzing gaps in the process. The "alpha-algorithm" plug-in within the process mining software is used to visualize a Petri Net graph and convert it to a BPMN diagram, providing a process model that summarizes the sequences of activities followed by most/all cases in the log.	BPM database and estimated KPI files	[13]

Appendix C
Quality Assessment

Study	Quality Criteria 1	Quality Criteria 2	Quality Criteria 3	Quality Criteria 4	Total Score
[13]	0	0	0	1	1
[40]	0	1	0	0	1
[18]	0	0	1	0	1
[22]	0	1	0	0	1
[2]	0	1	0	0	1
[31]	0	0	0	1	1
[9]	1	1	1	1	4
[11]	0	0	0	1	1
[8]	1	1	1	1	4
[28]	1	1	0	1	3
[27]	1	0	1	1	3
[42]	1	1	0	0	2
[7]	0	1	1	0	2
[21]	0	0	0	1	1
[38]	0	0	0	1	1
[37]	0	1	0	0	1
[5]	1	0	0	1	2
[43]	0	0	1	0	1
[16]	1	1	0	0	2
[12]	0	1	0	0	1
[19]	0	1	0	0	1
[32]	0	1	0	0	1

References

[1] W. M. P. van der Aalst et al. "Prom: The process mining toolkit". In: *Proceedings of the Business Process Management Workshops*. Vol. 1. Springer, 2009, pp. 11–16.

[2] Albakary; Sedky; Abdelmoez. "Interactive Bottleneck Detection in Data Driven Business Process Simulation in Healthcare: Egyptian Case Study". In: 1 (2022), pp. 234–240. DOI: 10.1109/ICCTA58027.2022.10206105.

[3] Víctor Augusto and Xiaolan Xie. "A literature review of the application of process mining in healthcare". In: *IEEE Access* 8 (2020), pp. 18382–18400.

[4] David I. Ben-Tovim et al. "Lean thinking across a hospital: redesigning care at the Flinders Medical Centre". In: *Australian Health Review* 31.1 (2007), pp. 10–15.

[5] Cho M; Song M; Park J; Yeom SR; Wang IJ; Choi BK. "Process Mining-Supported Emergency Room Process Performance Indicators." In: 1 (2020). DOI: 10.3390/ijerph17176290.

[6] C. de Blok et al. "Interfaces in service modularity: A typology developed in modular health care provision". In: *Journal of Operations Management* 31.7-8 (2013), pp. 345–361.

[7] Andrews; Wynn; Vallmuur; Ter Hofstede; Bosley. "A comparative process mining analysis of road trauma patient pathways". In: 17.1 (2020). DOI: 10.3390/ijerph17103426.

[8] Rojas; Cifuentes; Burattin.; Munoz-Gama; Sepúlveda; Capurro. "Performance analysis of emergency room episodes through process mining". In: 16.1 (2019). DOI: 10.3390/ijerph16071274.

[9] Rojas; Munoz-Gama; Sepúlveda; Capurro. "Process mining in healthcare: A literature review". In: 61.1 (2016), pp. 224–236. DOI: 10.1016/j.jbi.2016.04.007.

[10] Mark R. Chassin and Jerod M. Loeb. "High-Reliability Health Care: Getting There from Here". In: *The Milbank Quarterly* 91.3 (2013), pp. 459–490. DOI: 10.1001/jama.2013.277053.

[11] Arias M; Roja E; Aguirre S; Cornejo F; Munoz-Gama; Sepulveda; Capurro D. "Defining Healthcare KPIs Using Process Mining and Patient Journey Maps". In: 1 (2023). DOI: 10.1109/CLEI60451.2023.10345781.

[12] Erhard A; Arthofer K; Helm E. "Extending a Data Management Maturity Model for Process Mining in Healthcare." In: 301.1 (2023), pp. 192–197. DOI: 10.3233/SHTI230038.

[13] Elhadjamor; Ghannouchi. "Analyze in Depth Health Care Business Process and Key Performance Indicators using Process Mining". In: 164.1 (2019), pp. 610–617. DOI: 10.1016/j.procs.2019.12.227.

[14] Cynthia Haraden and Roger Resar. "Patient Flow in Hospitals: Understanding and Controlling It Better". In: *Nursing Administration Quarterly* 28.3 (2004), pp. 243–253. DOI: 10.1097/01.NNA.0000135148.60745.15.

[15] Wallace J. Hopp and Mark L. Spearman. *Factory Physics*. 3rd. Waveland Press, 2011.

[16] van Hulzen G; Martin N; Depaire B; Souverijns G. "Supporting capacity management decisions in healthcare using data-driven process simulation." In: 129.1 (2022), p. 104060. DOI: 10.1016/j.jbi.2022.104060.

[17] Alistair E. W. Johnson et al. *MIMIC-IV (version 0.4)*. https://doi.org/10.13026/s6n6-xd98. PhysioNet. 2021.

[18] Kurniati; Hall; Hogg; Johnson. "Process mining in oncology using the MIMIC-III dataset". In: 971.1 (2018). DOI: 10.1088/1742-6596/971/1/012008.

[19] Park K; Cho M; Song M; Yoo S; Baek H; Kim S; Kim K. "Exploring the potential of OMOP common data model for process mining in healthcare." In: 18 (2023), e0279641. DOI: 10.1371/journal.pone.0279641.

[20] B.A. Kitchenham, D. Budgen, and P. Brereton. *Evidence-Based Software Engineering and Systematic Reviews*. 4th ed. CRC Press, 2015.

[21] Ireddy; Kovalchuk. "An Experimental Outlook on Quality Metrics for Process Modelling: A Systematic Review and Meta Analysis". In: 16.1 (2023). DOI: 10.3390/a16060295.

[22] Helm; Lin; Baumgartner; Lin; Küng. "Towards the use of standardized terms in clinical case studies for process mining in healthcare". In: 17.1 (2020). DOI: 10.3390/ijerph17041348.

[23] Kurniati. "The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database". In: 25.1 (2019), pp. 1878–1893. DOI: 10.1177/1460458218810760.

[24] R. Lenz and M. Reichert. "IT support for healthcare processes–Premises, challenges, perspectives". In: *Data & Knowledge Engineering* 61.1 (2007), pp. 39–58.

[25] Jeffrey K. Liker. *The Toyota Way: 14 Management Principles from the World's Greatest Manufacturer*. McGraw-Hill Education, 2004.

[26] Eugene Litvak and Michael C. Long. "Cost and quality under managed care: irreconcilable differences?" In: *American Journal of Managed Care* 6.3 (2000), pp. 305–312.

[27] Rashed; El-Attar; Abdelminaam; Abdelfatah M. "Analysis the patients' careflows using process mining". In: 18.1 (2023). DOI: 10.1371/journal.pone.0281836.

[28] Pereira; Santos; Maceno. "Process mining project methodology in healthcare: a case study in a tertiary hospital". In: 9.1 (2020). DOI: 10.1007/s13721-020-00227-w.

[29] R. Mans et al. "Application of process mining in healthcare–A case study in a Dutch hospital". In: *International Journal of Health Information Management Research* 5.2 (2015), pp. 79–94.

[30] De Roock E; Martin N. "Process mining in healthcare - An updated perspective on the state of the art." In: 127.1 (2022), p. 103995. DOI: 10.1016/j.jbi.2022.103995.

[31] Marshall; Novakovic. “Process Mining the Performance of a Real-Time Healthcare 4.0 Systems Using Conditional Survival Models”. In: 15.1 (2022). DOI: 10.3390/a15060196.

[32] Kurniati AP; Rojas E; Hogg D; Hall G; Johnson OA. “The assessment of data quality issues for process mining in healthcare using Medical Information Mart for Intensive Care III, a freely available e-health record database.” In: 25.1 (2019), pp. 1878–1893. DOI: 10.1177/1460458218810760.

[33] Observational Health Data Sciences and Informatics (OHDSI). *OHDSI Athena: Standardized Vocabularies*. https://athena.ohdsi.org/. Accessed: 2024-09-04. 2024.

[34] J. Marc Overhage et al. “Validation of a common data model for active safety surveillance research”. In: *Journal of the American Medical Informatics Association* 19.1 (2012), pp. 54–60. DOI: 10.1136/amiajnl-2011-000376.

[35] Andrew A Partington et al. “Process mining for clinical processes: A comparative analysis of four Australian hospitals”. In: *ACM Transactions on Management Information Systems (TMIS)* 5.4 (2015), p. 19.

[36] E. Rojas et al. “Process Mining in Healthcare: A Literature Review”. In: *Journal of Biomedical Informatics* 61 (2016), pp. 224–236.

[37] Andrews R; Wynn MT; Vallmuur K; Ter Hofstede AHM; Bosley E; Elcock M; Rashford S. “Leveraging Data Quality to Better Prepare for Process Mining: An Approach Illustrated Through Analysing Road Trauma Pre-Hospital Retrieval and Transport Processes in Queensland.” In: 16.1 (2019). DOI: 10.3390/ijerph16071138.

[38] Sharma S; Srivastava S. “QCPW: A quality centric process workflow improvement approach for a legacy healthcare information system”. In: 14.1 (2020), pp. 129–137. DOI: 10.1049/iet-sen.2019.0064.

[39] Christoph Schröer, Felix Kruse, and Jorge Marx Gómez. “A Systematic Literature Review on Applying CRISP-DM Process Model”. In: *Procedia Computer Science* 181 (2021), pp. 526–534. DOI: 10.1016/j.procs.2021.01.199. URL: https://www.sciencedirect.com/science/article/pii/S1877050921002416.

[40] Rojas; Arias; Sepúlveda. “Clinical processes and its data, what can we do with them?” In: 1 (2015), pp. 642–647. DOI: 10.5220/0005287206420647.

[41] John S. Toussaint and Leonard L. Berry. “The Promise of Lean in Healthcare”. In: *Frontiers of Health Services Management* (2013).

[42] Gatta; Vallati M; Lenkowicz J; Casa C; Cellini ; Damiani A; Valentini V. “A framework for event log generation and knowledge representation for process mining in healthcare”. In: 2018-November.1 (2018), pp. 647–654. DOI: 10.1109/ICTAI.2018.00103.

[43] Jonk J; Schaller M; Netzer M; Pfeifer B; Ammenwerth E; Hackl W. “Process Mining of Nursing Routine Data: Cool, but also Useful?” In: 293.1 (2022), pp. 137–144. DOI: 10.3233/SHTI220360.

[44] James P. Womack, Daniel T. Jones, and Daniel Roos. *The Machine That Changed the World: The Story of Lean Production*. Free Press, 1990.