Utilizing the qSOFA-score & process mining techniques to predict ICU admissions

Written by: **Jelle Gerritsen**
Assigned and supervised by: **Faiza Bukhsh**
Study, institution: **Creative Technology, Twente University**
Assignment type: **Bachelor thesis**
Date: **29-04-2024**
Front page image by: **Dreamshaper 8 (2024)**

Abstract

Objective: The graduation project aimed to investigate patient event logs of patients going into the ICU and generate an ICU admission predictor using the qSOFA score and process mining techniques.

Methods: Data was simulated with qSOFA scores that were linked to patient outcomes and the data was attached to a patient journey event log dataset of general practitioner patients. A lab test in the patient journey dataset was used as a substitute for the ICU. After combining the datasets, they were analysed using the "Perform Predictions of Business Process Features" plugin from Prom Lite 1.4 to analyse the results.

Results: 10000 patients were in the dataset of which 2106 went to the ICU. With the qSOFA-scores attached, it was concluded that qSOFA ≥ 2 was the most accurate predictor with a 73.8% accuracy which is not in line with background research. Looking at two control metrics the kappa statistic and the RMSE which were 0.2379 and 90.8% accordingly, the metrics both suggest that the predictor cannot be taken as scientifically valid as the kappa statistic should be in the range of 0.81-1.00 and the RSSE which should be between 0%-10%.

Conclusion: Even though the resulting ICU admission predictor, qSOFA ≥ 2, had a strong accuracy, it was not in line with background research and cannot be taken as scientifically valid. The simulated data is highly likely to not be an accurate representation of real qSOFA-to-ICU outcome data. However, utilizing process mining techniques on real data could still be beneficial to the ICU management field.

Table of content

1. Introduction

ICU admissions are highly costly in terms of money and resources. A study by Dasta et al. (2005) tried to quantify the costs of a single ICU admission and found that the mean ICU admission cost was 19,725 dollars. Only in the Netherlands there were 64.868 ICU admissions in 2020 according to Stichting NICE (2023) on their "data in beeld" page. Multiplying those numbers would give a significant 1,279,521,300 dollars that is spent on ICU admissions in the Netherlands only. 67.2% of all ICU admissions were unplanned ICU admissions.

There are multiple potential causes for an ICU admission which can be put into five primary categories according to the Society of Critical Care Medicine (n.d.) which are "respiratory insufficiency/failure with ventilator support, acute myocardial infarction, intracranial hemorrhage or cerebral infarction, percutaneous cardiovascular procedures, and septicemia or severe sepsis without mechanical ventilation". However, the same article informs us that "Other conditions and procedures involving high ICU use are poisoning and toxic effects of drugs, pulmonary edema and respiratory failure, heart failure and shock, cardiac arrhythmia and conduction disorders, renal failure with major complication or comorbidity, gastrointestinal hemorrhage with complication or comorbidity, and diabetes with complication or comorbidity.". The fact a high variety of medical diagnoses, procedures and/or conditions could cause an ICU admission makes it challenging to predict an ICU admission. However, it could be highly beneficial if ICU admissions could be predicted as predicting an ICU admission could help preventing the progression of a medical condition and in some cases preventing an ICU admission by swift medical interventions. Preventing progressions of medical conditions will help to reduce costs for hospitals and patients as advanced surgeries might be prevented. It should not be forgotten that preventing progressions and early treatment will also improve the health outcomes of patients and reduce their mortality rates when progressions of medical conditions can be prevented. Lastly predicting ICU admissions allows for better ICU and surgery planning as it will be easier to understand the patient flows and therefore the anticipated needs for ICU beds. As this thesis report states earlier, 67.2% of all ICU admissions are unplanned. Even though they are unplanned it might be possible to predict a majority of those admissions that would come with the above-stated benefits.

In the past many researchers put effort into trying to find factors that could indicate an ICU admission, (Bentrem et al., 2005; Holguín et al., 2008; Chioncel et al., 2012; Monaco et al., 2019; Ju et al., 2021). However, a new scientific research field is on the rise which is process mining. Process mining is a technology field that utilizes data science for investigating and optimizing processes. The patient care process can also be dissected which can hold many new opportunities, especially when it comes to predicting ICU admissions. Utilizing process mining technologies for predicting ICU admissions is especially interesting because currently more and more data is collected in hospitals which would mean the solution created in this graduation project could be utilized all over the world.

Most hospitals have technologies that allow them to measure basic biomarkers which are biological variables that could be combined and utilized to predict or measure certain medical conditions. Some biomarkers could also be used to predict ICU admissions. The qSOFA score is a well-validated biomarker that is known to potentially have predictive abilities when it comes to ICU admissions. Other than being scientifically validated, the qSOFA score has another advantage. Measuring the qSOFA score is not very complicated and therefore even more basic hospitals should have the equipment available to measure it.

The goal of this graduation project is to build a model that can predict future ICU admissions utilizing process mining technologies and the SOFA score. The ideal outcome of this project would be that ICU admissions are detected earlier which could improve clinical outcomes for patients and reduce the amount of money and resources that are spent on ICUs. Hospitals could then spend those freed resources on other or potentially advanced healthcare which could make room for new opportunities to treat people better and overall healthcare improvement.

In this graduation project utilizing process mining technologies for predicting ICU Admissions using the qSOFA score is investigated. The main research question therefore is: "Can an ICU admission predictor be generated using process mining techniques and qSOFA Data?".

This thesis paper is built up in the following order. If the reader reads the paper from top to bottom, the reader has already finished the abstract and the current chapter, chapter 1 which is the introduction. From here, background research is done first which is chapter 2. Chapter 2 is also the chapter where the background research questions which are subquestions 1 to 4 are answered. This is followed by chapter 3 where the research structure is determined. Then in chapter 4, the steps that have been performed to generate the results are shown. In chapter 5 the results are analysed, and the validation research questions are answered which are subquestions 5 and 6. Then in chapter 6 the limitations and imperfections of this project are discussed. In chapter 7, the concluding remarks are made and some advice on future work is written down. Lastly, only the references and the appendices remain for extra information.

2. Background research

In this chapter, all background research was done. Firstly, four background research subquestions were formulated:

Subquestion 1: What are the current best ways to predict ICU admissions?
Subquestion 2: How are ICU admissions predicted in practice?
Subquestion 3: How is the qSOFA score related to ICU admissions?
Subquestion 4: How can process mining be used for making ICU admission predictions?

In section 2.1 the state of the art was briefly discussed as well as other existing alternatives that could help with predicting ICU admissions which answers research subquestions 1 and 2. In section 2.2 more background research on the SOFA & qSOFA was done such that the scores could be understood and the predictive abilities for ICU admissions could be estimated which answers research sub question 3. Lastly, in section 2.3 the answer to sub question 4 on how process mining can help with making better predictions was formulated.
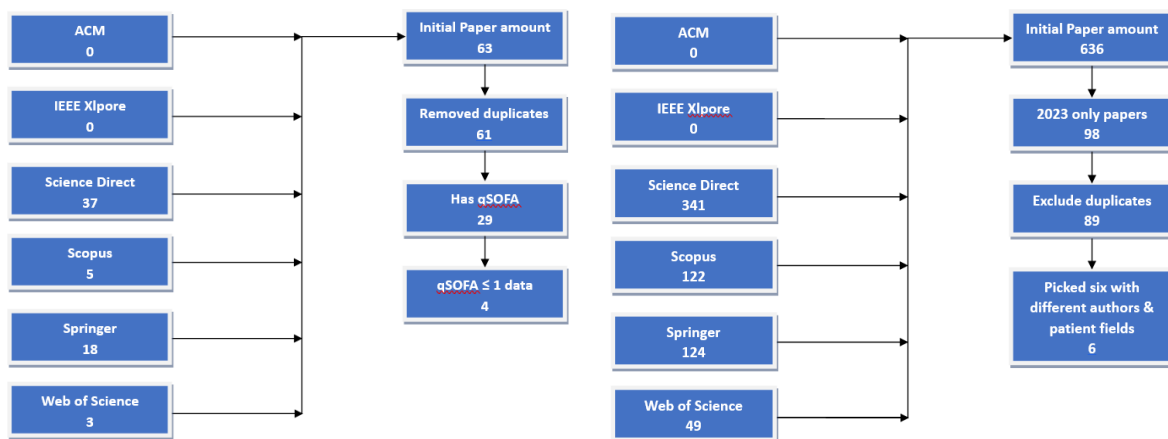

Figure 1: Study selection process for subquestion 1 (left) and subquestion 3 (Right)

Different methods were used to answer the literature research subquestions. The search query method, as described by Kitchenham et al. (2009), was used to answer subquestions 1 and 3. To answer subquestion 1, the query "Predictor of ICU admission" was used. This gave a total of 636 results as predictors of ICU admissions are studied a lot. To describe the significant disagreement between researchers and prevent the article from becoming the size of a book, the decision was made to limit the search for exclusively 2023 articles such that the most recent literature was used to describe the state of the art. There were still 98 articles left to review. After taking out duplicates 89 articles remained. The decision was made to choose six articles that were written by different authors and that they were in slightly different niches to paint a picture of the state of the art. This was done because it was concluded that even though much research was done, there is still too little agreeance on which factors are relevant enough and which are not. Subquestion 2 is answered as a reaction to the literature from subquestion 1 combined with some basic knowledge about hospitals. For subquestion 3 the query: ""Sequential Organ Failure Assessment" AND "Predictor of ICU Admission"" was used. 63 papers were found. After the decision was made that the qSOFA-score was used instead of the original SOFA score there were two hypotheses left. qSOFA ≥ 1 and qSOFA ≥ 2. Both hypotheses had 4 papers with data to back them up. In the end, qSOFA ≥ 1 was chosen, because after reviewing the data of the papers qSOFA ≥ 2 had a 65% accuracy on average

and qSOFA ≥ 1 had a 72% accuracy on average. Lastly, to answer subquestion 4, the book "Process Mining: Data Science in Action" (Van der Aalst, 2016) was used combined with the knowledge obtained from Prom Lite 1.4.

2.1 Predicting ICU admissions

Many different symptoms or measurements could serve as an indicator of a patient's need for ICU care. Mahammadova et al. (2023) did a study on patients who underwent scoliosis surgery and found that predictive factors are having restrictive lung disease, epilepsy, neuromuscular scoliosis, a high fusion level, the need for transfusion, a long operative duration, low Hb and pH values, and high lactate values. Cresti et al. (2023) examined features that were associated with a higher risk on ICU admissions in patients with infective endocarditis. They found that mechanical ventilation, SOFA score > 5, GCS ≤ 8, renal insufficiency, septic shock, cerebral embolism, and surgical indication are significant risk factors. Feghoul et al. (2023) did a study determining risk factors for patients with SARS-CoV-2 in which they identified diabetes, obesity, hepatitis, fever, dyspnoea, oxygen requirement, and TTV load as predictors of ICU admission. Naser et al. (2023) worked on finding significant factors in patients with COPD and found that aging 65 and older, smoking, receiving home care, and having specific comorbidities which in the paper are summarised by: "hypertension, Diabetes mellitus, Ischemic heart disease, Obstructive sleep apnea syndrome, heart failure, stroke, and Tumour or malignancy" are significant factors. Zahran et al. (2023) also researched many factors but only found high respiratory rate (≥ 22) and recent chemotherapy to be significant predictors of ICU Admissions. Qin et al. (2023) did research on patients with autoimmune encephalitis and found hypoventilation and increased NLR (Neutrophil-to-lymphocyte ratio) to be indicators of increased ICU admission risk.

Even though, many studies have been done on factors that influence the risk on ICU admissions science is not yet completely in agreeance about which factors have a significant influence. Most problems arise from the fact that there are many different potential causes for an ICU admission. Also, many studies that are done on the topic seem to always specify a specific group that is studied which makes it harder to distinguish between factors that pose a universal risk versus factors that indicate risk given that the patient already has a certain condition. This, however, does imply that warning scores and scoring systems, like the SOFA score, might be the key to universally assessing the risk of an ICU admission as they take in multiple variables that combine into a more universally effective measurement.

There are multiple ways to predict ICU admissions, however, in practice clinical expertise is still the norm. Doctors and therapists are responsible for what care patients get and under normal circumstances, no measurement tool checks any variables that would indicate a later ICU admission. The reason for this is mainly that scientific literature is still in strong disagreement and proper ways to accurately predict ICU admissions have not been invented yet. Certain methods are quite accurate and in section 2.4 it is found that the qSOFA has a 72% accuracy which is promising. However, this is still too low to provide consistent value in hospitals.

In hospitals, doctors make decisions on how significant the risk on an ICU admission is and whether certain actions need to be taken. Certain patients might need certain monitors attached to them; others might not need any form of monitoring. This is, however, completely dependent on the judgment of the care-taking doctor. Of course, the monitors might also help alert doctors to potential risks, but monitoring is not a standard procedure. Also, certain

surgeries have a significant risk of landing someone in the ICU. However, doctors are mostly solely responsible for assessing the risk. This is, again, mainly caused by the fact that science does not yet agree on how to accurately predict ICU admissions. Another significant argument for relying mainly on the clinical expertise of doctors is that society right now prefers that a human is responsible for making critical decisions in intensive healthcare rather than technology.

The challenge with clinical expertise is that it might be time-intensive and potentially challenging to assess the risk of an ICU admission. An example would be that a patient comes in with a broken bone after a crash, but no other signs of danger. In this scenario, a doctor could potentially find no other things to investigate, but an objective measurement tool might help finding hidden issues. Assisting doctors in risk assessment could help saving time and resources and could help spotting hidden risks which could lead to earlier treatment and improve patient outcomes which should be the number one priority.

2.2 The SOFA & qSOFA score

The SOFA score is a biochemical assessment tool created objectively describing the degree of organ failure over time in individual patients and groups of patients with sepsis and was designed in 1994 during a consensus conference organized by the European Society of Intensive Care and Emergency Medicine (Moreno & Metnitz, 2008) and published in a paper by Vincent et al. (1996). The SOFA score takes 6 separate systems into account: the central nervous system measured with the Glasgow Coma Scale (CGS), the Cardiovascular system measured via Hypotension, the respiratory system measured with PaO_2/FiO_2, the liver measured with bilirubin mg/dL, the kidneys also known as the renal system measured with (Creatinine mg/dL) & blood clotting also known as coagulation (Platelet Count).

All systems can get a score between 0 and 4 which results in a maximum score of 4 * 6 = 24. When the SOFA score ≥ 2 then life-threatening organ dysfunction is predicted. For the central nervous system, the Glasgow coma scale is used each score on that scale is related to one of the scores 0-4. For the cardiovascular system hypotension is measured in mean arterial pressure (MAP). MAP is calculated using the following formula: $MAP = diastolic\ blood\ pressure + \frac{1}{3}(diastolic\ blood\ pressure - systolic\ blood\ pressure)$ (DeMers, 2023). If MAP < 70 millimetres of mercury (mm HG) the score is ≤ 1. If any dobutamine or dopamine ≤ 5 microgram per kilogram (patient weight) per minute (µg/kg/min) is used the score is 2. If noradrenaline ≤ 0.1 µg/kg/min is used or dopamine between 5-15 µg/kg/min is used the score is 3 and finally, if more than 0.1 µg/kg/min noradrenaline or more than 15 µg/kg/min dopamine is used the score is 4. For the respiratory system, the PaO_2/FiO_2 is calculated, also known as the P/F ratio. This is done by dividing the partial pressure of arterial oxygen (PaO_2) which is the concentration of the amount of oxygen in the arterial blood by the fraction of inspired oxygen, the amount of oxygen that is delivered to a patient. All measurements are related to one of the scores 0-4 for the sofa score. For the liver, the bilirubin concentrations in a patient's blood are measured in milligrams per decilitre of blood. For the renal system, the creatinine concentrations in a patient's blood are measured in milligrams per decilitre of blood. All concentrations measured are related to one of the scores 0-4. Lastly, for blood clotting/coagulation, the number of platelets in the patient's blood is measured in 10^3 platelets per microliter of blood. Table 1 is a summary of all discussed criteria.

System	0	1	2	3	4
CNS (Glasgow Coma Scale)	15	13-14	10-12	6-9	6 <
CVS (Hypotension)	No hypotension	MAP < 70 mm Hg	any dobutamine **or** dopamine ≤ 5 µg/kg/min	NA ≤ 0.1 µg/kg/min **or** dopamine 5-15 µg/kg/min	NA > 0.1 µg/kg/min **or** dopamine 15> µg/kg/min
Resp system (PaO$_2$/FiO$_2$)	> 400	300-400	200-300	100-200	< 100
Liver (bilirubin mg/dL)	< 1.2	1.2- 1.9	2.0-5.9	6.0-11.9	> 12.0
Renal (Creatinine mg/dL)	< 1.2	1.2-1.9	2.0-3.4	3.5-4.9	> 5.0
Coagulation (Platelets 10^3/ mm^3)	>150	100-150	50-100	20-50	<20

Table 1: Criteria for SOFA score system (Vincent et al., 1996).

After further background research, the qSOFA (quick Sequential Organ Failure Assessment) score was encountered often. Significantly more often than the normal SOFA score. The qSOFA score is a score that is designed to be easier and quicker for drawing conclusions and is derived from the SOFA score. The qSOFA has three conditions: Glasgow Coma Scale < 15, Systolic blood pressure ≤ 100 mm Hg & Respiratory rate ≥ 22 breaths per minute. If two or more conditions are true, qSOFA is positive. Even though qSOFA and SOFA are not in a 1-to-1 relation they are comparable. After finding that there is more scientific backing for the qSOFA score in relation to ICU admissions and after finding that its relation to ICU admissions is more accurate than SOFA ≥ 2. It was decided to continue solely with the qSOFA score. In table 2 an overview of the qSOFA scoring system can be found.

System	0	1
CNS - Impaired consciousness (Glasgow Coma Scale)	15	< 15
Resp system - Respiratory rate (Breaths per minute)	< 22	≥ 22
Artery system - Systolic blood pressure (millimetres of mercury)	> 100	≤ 100

Table 2: Criteria for qSOFA score system in table form Seymour et al. (2016)

2.3 The qSOFA score related to ICU admissions

Secondly, most studies had different ways to analyse the accuracy of the qSOFA, however almost all of them calculated the sensitivity and specificity. In the background research accuracy is defined as the average of the combined sensitivity and specificity. In this case, sensitivity is how often a model correctly diagnoses an individual with a condition as positive

and specificity is how often a system correctly diagnoses an individual without a condition as negative. Sometimes papers only calculated the OR (Odds ratio) which displays a likelihood of scenario A given scenario B or the AUC (Area under the curve) which uses a curve and tries to determine how much of the area is under the assumed correct curve. AUC is also a metric of accuracy, but not all studies had that data and therefore the combined average of the sensitivity and specificity was used.

Four studies were found that examine the predictive abilities of the qSOFA score. Zhang et al. (2020) found a 68% accuracy for the qSOFA score ≥ 1 with a sensitivity of 70.11% and a specificity of 65.32%. Bae et al. (2022) found results that might be even more promising as their found accuracy was 75% with a sensitivity of 71.3% and a specificity of 79.6%. Chu et al. (2020) found even higher results as their qSOFA ≥ 1 gave a 79% accuracy with a sensitivity of 82.1% and a specificity of 76.6%. Lastly, the study of Covino et al. (2020) was examined and interestingly they found a significantly lower sensitivity of 53.8%, but still the highest specificity of 80.2% resulting in an accuracy of 67%.

It might be worth noting that Covino et al. had a significantly smaller sample size of 334 samples followed by Zhang et al. with 742 samples, then Bae et al. with 1151 samples, and finally the largest study was the one by Chu et al. with 3561 samples.

To summarize, on average the accuracy of the qSOFA score ≥ 1 was 72% if calculated by combining and taking the averages of the sensitivities and specificities. The average sensitivity was 69% and the average specificity was 75%. These averages are not weighted by sample size and do equally contribute to these final numbers.

2.4 Making predictions using process mining

Process mining is a set of techniques that might improve predictive abilities for any event. The main advantage of Process mining is being able to track down all previous events before the event that needs to be predicted. Being able to track the history of an individual that undergoes a certain process is valuable.

Even though there was no single inventor of process mining. Wil van der Aalst has been a key figure. In his book "Process Mining: Data Science in Action" he describes process mining like this: "Process mining bridges the gap between traditional model-based process analysis (e.g., simulation and other business process management techniques) and data-centric analysis techniques such as machine learning and data mining" (Van der Aalst, 2016). Van der Aalst writes that process mining allows for new strategies to improve processes in multiple domains.

In chapter 10 of Van der Aalst's book (Van de Aalst, 2016) real-time operational support is discussed. Van der Aalst proposes regression analysis or decision tree learning for predicting a feature, in our case an ICU admission.

To make predictions most process mining software is supplied with algorithms to build prediction models. The software used for this project is Prom Lite 1.4. This program has a plugin named: "Perform Predictions of Business Process Features" which can generate results by selecting the variables that need to be predicted and creating an augmented event log. Then in the other tab, the dependent variable, and the type of tree it needs to generate can be selected. Screenshots of the plugin are visible in figure 2.
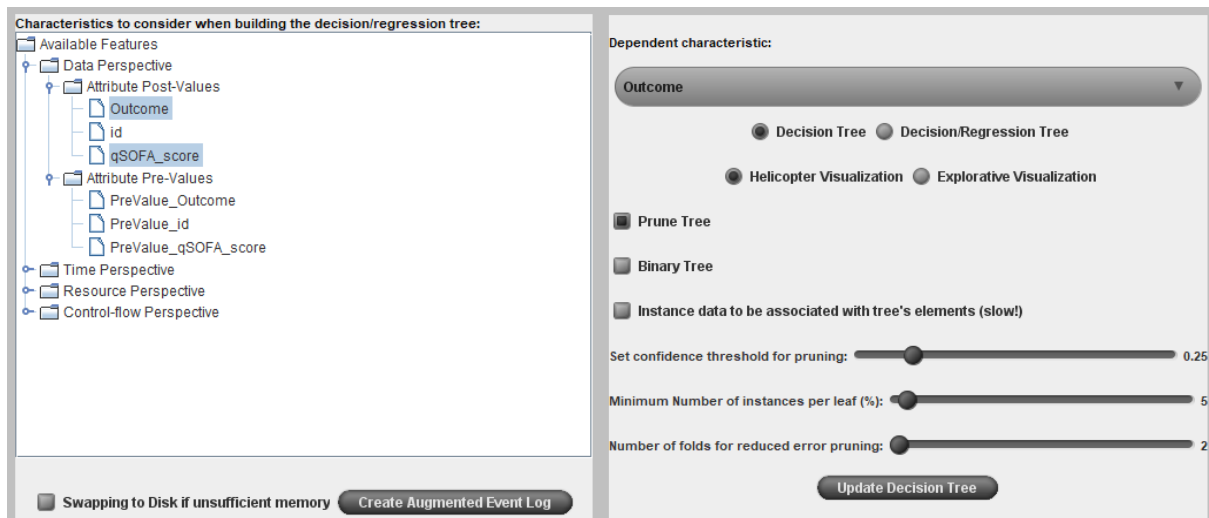
Figure 2 – Doing predictions in Prom Lite 1.4

Summary

In this chapter literature research was done to find answers to subquestions 1 to 4 and lay the foundation for the methods and choices in this research project. Firstly, it was found that many studies investigate the predictors for ICU admissions, but that studies do not clearly agree on what factors play a role and what factors are disease-specific. Also, in practice, no methods are used on a consistent basis in hospitals to predict ICU admissions. In this chapter, it was also found that the qSOFA score has some significant ability to predict ICU admissions. Lastly, it was found that process mining is a set of techniques with significant potential in scientific fields where a process has a consistent structure such that analysis over time is valuable.

3. Methodology

In this chapter, the six phases of the PM^2 methodology by van Eck et al. (2015) were explained which gave a basic direction and structure to this thesis. The planning phase is described in chapter 2, background research, and this chapter (Chapter 3, thesis structure). The phases 2, 3 & 4 are written out in chapter 4, realization. Chapter 5 "Analysis" describes the fifth phase and the last chapters represent phase 6. It should, however, be noted that actual improvement of the process and support is outside the scope of this thesis, but recommendations for process improvement and future research are written in those chapters.

3.1 Research structure

There are multiple approaches for structuring a process mining research project. After thorough consideration, the PM^2 method by van Eck et al. (2015) was chosen. The advantages of PM^2 over other methods where the methodology is developed specifically for process mining, the methodology is highly iterative which fits with the Creative Technology design process & the methodology is designed to be applicable for non-structured processes like the patient flow inside an ICU or hospitals. In figure 3 an overview of PM^2 can be found.
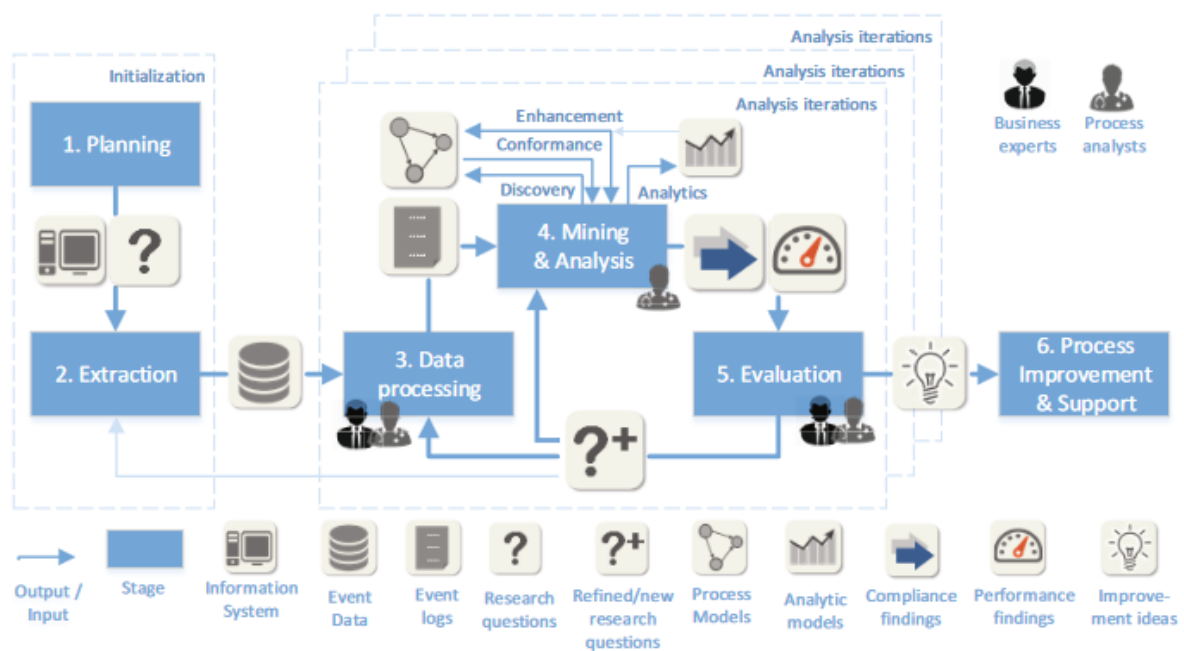

Figure 3 - An overview of the PM2 methodology

1. Planning phase

The purpose of the planning phase is to set up the project. Selecting a process to study and defining the goal of the project are both important. After this, identifying the main research question and sub questions is done. The planning phase is mainly written out in the chapter 2 "Background Research" and this chapter "Methodology".

2. Extraction phase

The purpose of the extraction phase is to acquire the necessary data and the tools needed to process, mine & analyse the data. Also, the scope of the data should be determined which for process mining means what exact data with what time frame needs to be collected. Doing background research and context analysis is also part of this phase as this phase requires knowledge about the process. This phase is mainly written out in section 4.1 and technically also in sections 4.2 and 4.3. Originally the plan was to extract the data that was obtained in sections 4.2 & 4.3. However, as extraction from the MST hospital was no longer possible, the data was designed from scratch and combined with a general practitioner patient journey dataset which was used to generate results instead.

3. Data processing

The purpose of data processing is to process the data in such a way that it can be mined and analysed. In this phase, the data is filtered and prepared. Sometimes the data can also be aggregated and enriched. Data aggregation means that two or more events are merged into one which can allow more efficient process mining. When enriching data, extra attributes are added to the event logs. Enriching data also allows more efficient process mining and can sometimes create new insights. Enriching data can be done by specifying the event logs themselves or by adding additional information from external data. In section 4.4 the process for data processing is written out.

4. Mining & Analysis

The mining & analysis phase is the phase where process mining techniques are applied to the extracted and processed data. In this phase, there are four main activities: process discovery, conformance checking, enhancement, and process analytics. Process discovery is the first phase where the first process models are made to get an idea of the process. After discovery, the process model should be checked for conformance. In this part, the model is checked on multiple aspects e.g., time, quality, resources & cost to verify whether the model is displaying behaviour that is like or at least strongly comparable to the real process. Process enhancement follows conformance checking. The aim of process enhancement is to improve the existing process model using knowledge about the real process. Lastly, process analytics can be done. Process analysis techniques are techniques that understand processes using data mining techniques or visual analytics to gain more insight into the generated process model. In section 4.5, the mining and analysis process was described.

5. Evaluation phase

The evaluation phase aims to gather all findings in the mining & analysis phase and to evaluate those findings to set up ideas for improvement that help achieve the goals of the project. Two main activities are described for this stage: Diagnose and Verify & Validate. Diagnosing the findings from phase 4 includes: correctly interpreting the results, separating expected from unexpected results & adapting existing or identifying new research questions from future iterations. Verify & Validate is about investigating findings and figuring out how they differ from the real process. Verification is about comparing the original data to the data that the process

model produced. Validation, on the other hand, compares the claims of stakeholders via e.g., interviews, questionnaires, or feedback sessions to the findings. The evaluation phase is written out in the results, chapter 5. In this chapter, the model and the data are evaluated.

6. Process Improvement & Support

Process Improvement & Support is the last step and is done after multiple iterations of phases 3, 4 & 5. Process Improvement & Support is about implementing the knowledge gained in the research. This phase has two main activities: Implementing improvements & supporting operations. Implementing improvements is about applying the changes to the process as this is often the goal of the project. However, as the paper from van Eck et al. (2015) states: "The actual implementation of process modifications is generally a separate project and a different area of expertise. The results of a process mining project then form the fact-based input of such process improvement efforts." The paper does, however, mention techniques that focus on the implementation process of process improvement which are business process re-engineering and Six Sigma (Harmon, 2010). As mentioned, phase 6 is most often a separate project and is therefore outside the scope of this research project.

Summary

In this chapter, the research structure was laid down using the PM^2 methodology as the foundation. All six phases of the PM^2 methodology were described and directions were given on where each phase is written out in this thesis. Phase 6 "Process Improvement and support" of the PM^2 method was found to be the only phase that was outside the scope of this research project. However, recommendations for future work are made in chapter 6 "Discussion" which can be seen as a part of a version of phase 6.

4. Realization

In this chapter, the realization of the thesis was written out which relates to phases 2 and 3 of the PM^2 methodology. Firstly, in section 4.1 the struggle to find suitable patient journey data was written out. In section 4.2 the process to simulate qSOFA-scores in relation to ICU admission outcome was described. In section 4.3 the process to attach the simulated qSOFA data to the patient journey event logs was described. Finally, in section 4.4 the process of mining and analysis is described which describes how the results of chapter 5 are obtained. It must be noted that sections 4.1 and 4.2 went differently than was planned for. The original idea was to obtain a dataset from the MST hospital which were event logs from patients that went into the ICU. The idea was that the dataset also included SOFA or qSOFA scores of patients such that conclusions could be drawn on predicting ICU admissions. As the MST hospital was no longer able to supply us with data, it was necessary to get a dataset that had similar patient flow and the qSOFA scores needed to be simulated and attached to the patient journey dataset.

4.1 Patient journey dataset

The hardest part of the thesis was to find suitable data. In the beginning, the MST hospital of Enschede would provide the necessary data. However, they were not able to deliver the data in time. The second option was trying to get access to the MIMIC-III database. For this, it was necessary to follow a data ethics course named: "CITI Data or Specimens Only Research". After investing 3 weeks of time, the course was completed, and it was finally possible to become a credentialed PhysioNet user which is required for getting access to MIMIC-III. In figure 4, a summary of the training report which is obtained after completion of the course is shown, the full training report can be found in the appendix. However, the application for credentials was rejected multiple times without clear argumentation. After this another source was tried, the task force on process mining's BPI Challenges those datasets were not quite suitable for this project. Finally, after looking at a fourth potential source 4TU.ResearchData, a dataset was found that was suitable. The dataset is about the patient journey of ten thousand patients making appointments with their general practitioners and undergoing a treatment process. (4TU.ResearchData, 2023)

• Name:	Jelle Gerritsen (ID: 12850776)
• Institution Affiliation:	Massachusetts Institute of Technology Affiliates (ID: 1912)
• Institution Email:	j.w.s.gerritsen@student.utwente.nl
• Institution Unit:	Student
• Curriculum Group:	Human Research
• Course Learner Group:	Data or Specimens Only Research
• Stage:	Stage 1 - Basic Course
• Record ID:	59873965
• Completion Date:	19-Dec-2023
• Expiration Date:	19-Dec-2026
• Minimum Passing:	90
• Reported Score*:	92

Figure 4 – A summary of the training report of the "CITI Data or Specimens Only Research" course required for becoming a credentialed PhysioNet user

4.2 Process discovery

Before data could be simulated, the best strategy to let the patient journey dataset be a sufficient substitute for a real dataset needed to be explored. For this, the process of the patient journey dataset needed to be discovered. As in preparation, the Introduction to Process Mining with ProM by Verbeek & Buijs (n.d.) was completed, the first logical step in discovering the process was applying an alpha miner to the process. Prom Lite 1.4 allows for 4 versions of the alpha miner: "Alpha", "Alpha+", "Alpha++" & "Alpha#". In combination with the NXML legacy classifier that classifies the events all alpha miner versions were tried on the dataset and in figures 5, 6, 7 & 8, you can see the results.
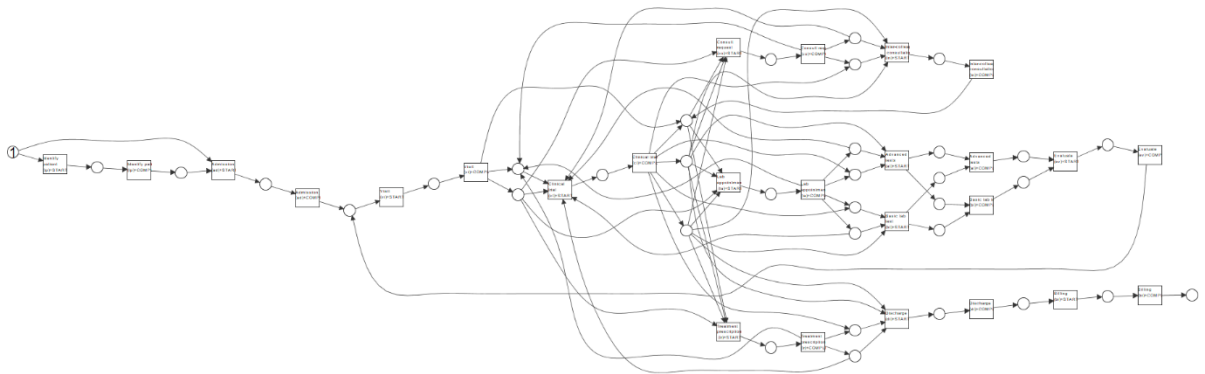
Figure 5 – Petri Net process model by alpha miner version "Alpha"
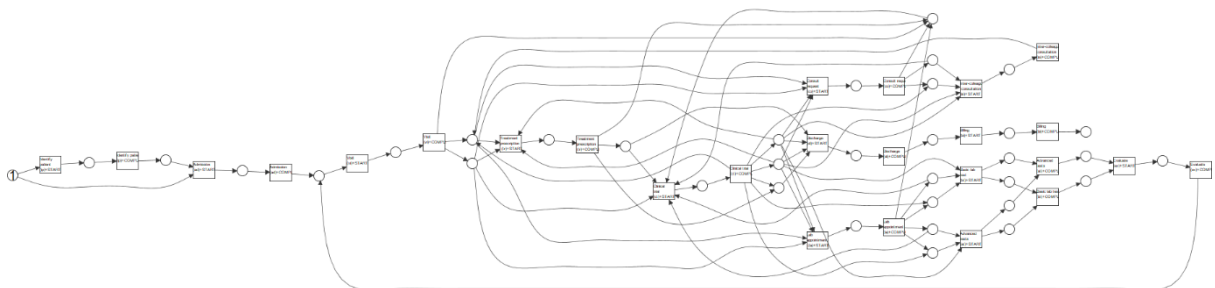
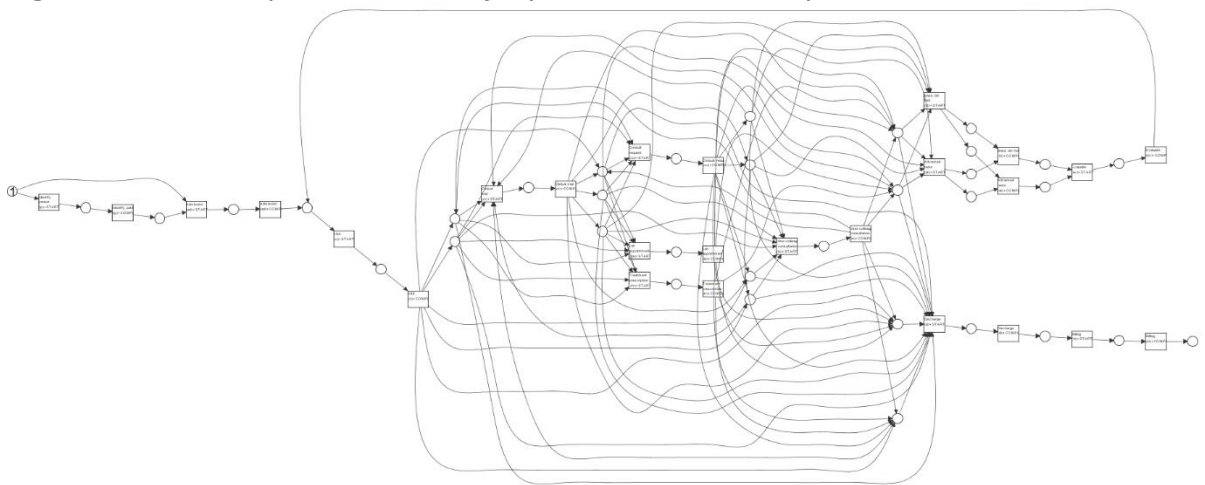*Figure 6 – Petri Net process model by alpha miner version "Alpha+"*

Figure 7 – Petri Net process model by alpha miner version "Alpha++"
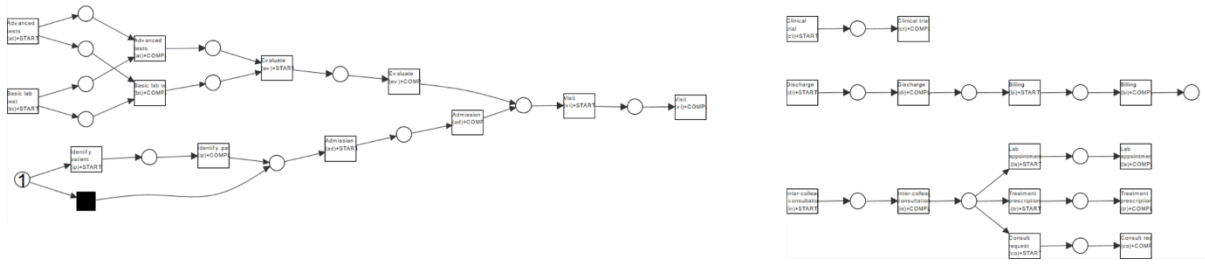
Figure 8 – Petri Net process model by alpha miner version "Alpha#"

Even though using the alpha miner provides us with some insights, the alpha miner does not guarantee a sound process model and none of the previous models were sound and encompassing the whole process. To ensure that the process model was sound an inductive miner with a 0.2 noise threshold and the NXML legacy classifier were used. The result is visible in Figure 9. The figure was sliced in two and stretched to slightly improve visibility.
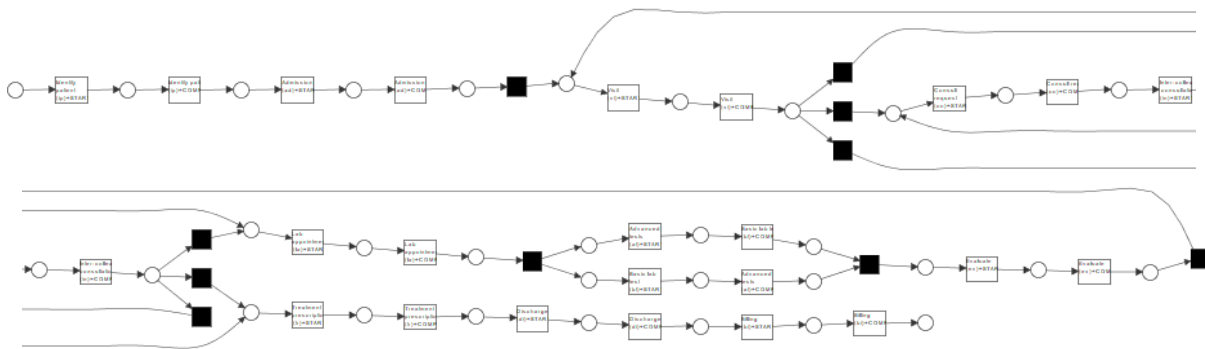

Figure 9 - Petri Net process model by an inductive miner with a 0.2 noise threshold
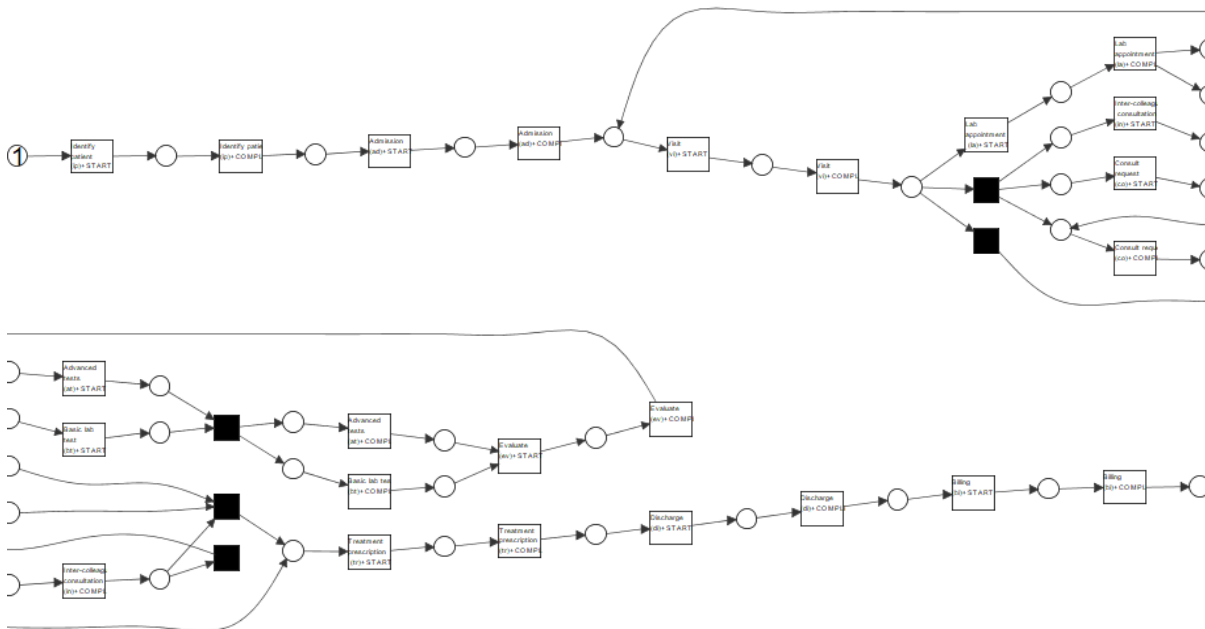

Figure 10 – Petri Net process model by the iD-heuristics miner on default settings

Even though this was a great improvement over the alpha miner algorithms the model was still slightly over-complicated. For example, the consult request start was put in parallel with the inter-colleague consultation start, even though a consultation request always precedes an

intercollege consultation. In pursuit of finding a more accurate model the interactive data-aware heuristics miner on default settings was used and gave a more promising result which is visible in figure 10. Also, for the heuristics miner the NXML legacy classifier was used the classify the events. The figure was sliced in two and stretched to slightly improve visibility.

The interactive data-aware heuristics miner gave the best and most simplified result while remaining significant log fitness. Now that the process model has been established a part could be chosen to substitute the ICU. This model does not include ICU admissions but is representative of a potential patient journey. So, an event that does not occur to all patients, but still to a significant amount of the patients was chosen to act like the ICU. For this project, the lab appointment event was chosen as it is occurring to 2106 out of 10000 patient journeys. This is perfect as it is irregular, but not so uncommon that noise will heavily influence results. After considering this, the actual dataset could be generated as the sample sizes of both distributions are now also known.

4.3 Data simulation

In this paper, a model is made that tries to predict whether a patient will end up in the ICU based on the qSOFA score. As no real qSOFA data could be acquired in the given timeframe for the thesis, simulated qSOFA scores were required. A log-normal distribution was used to represent the qSOFA scores. The patient journey dataset that is used has 10000 patients which is rather large. Therefore, it might be that a certain type of normality could be assumed. Knowledge about the qSOFA score also helps in choosing a distribution. The qSOFA-score is designed in such a way that a perfectly healthy individual should have a qSOFA-score of 0 and that having a higher qSOFA-score is less likely to occur. The chance that multiple parts of the body "malfunction" at the same time is less likely. A qSOFA-score of 3 is therefore most rare. This means that the distribution for the qSOFA-score is probably positively skewed. Lastly, taking the boundary qSOFA > 0, the most occurring number in the distribution, the mode, for patients that went to the ICU/had a positive outcome should be 1 and the mode for patients that did not go to the ICU/had a negative outcome should be 0. A distribution that could be shaped around the mode would be ideal for this project. The log-normal distribution, which is sometimes also referred to as Galton's distribution, after Francis Galton, satisfies all requirements and is commonly used in fields like science, medicine, and economics.

A program in python was written to generate two distributions: one distribution for qSOFA-scores of patients that went to the ICU/had a positive outcome and one for qSOFA-scores of patients that did not go to the ICU/had a negative outcome. Each distribution has a sample size, a standard deviation & a mode. The model was used to calculate the mean of the distribution, which is possible because a property of the log-normal distribution is that the $\text{Mode} = e^{(\mu - \sigma^2)}$ which can be rewritten to $\mu = e^{(\text{Mode} - 0.5\sigma^2)}$. The first try was with a standard deviation of one. Below is the result.

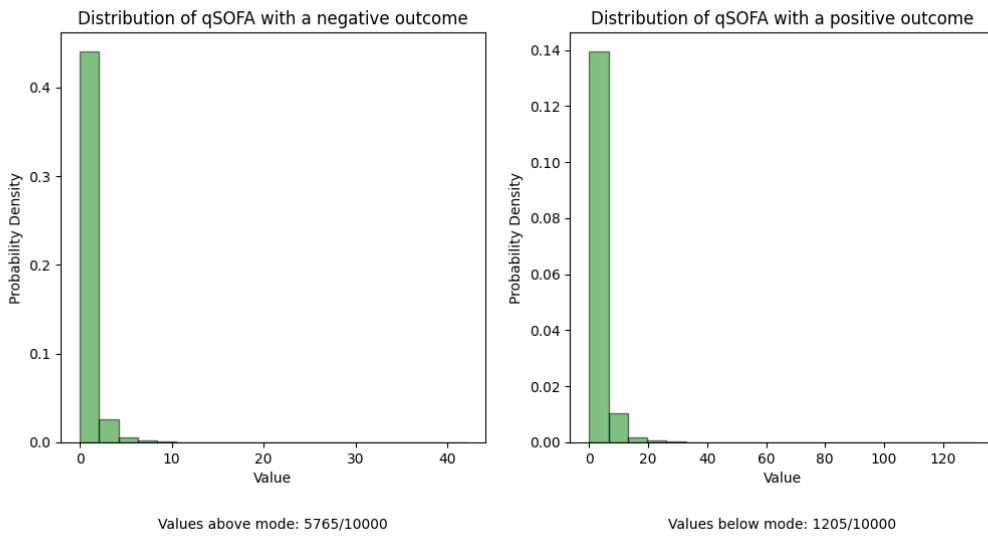Values above mode: 5765/10000   Values below mode: 1205/10000

Figure 11 – distribution without bounds and standard deviation = 1

This cannot be right as qSOFA does not go higher than three. To keep the distribution the same, but not allow for qSOFA > 3 or qSOFA < 0 all scores that were out of those bounds rerolled according to the same distribution until all scores lie in the range from 0 to 3. This generates the following result.
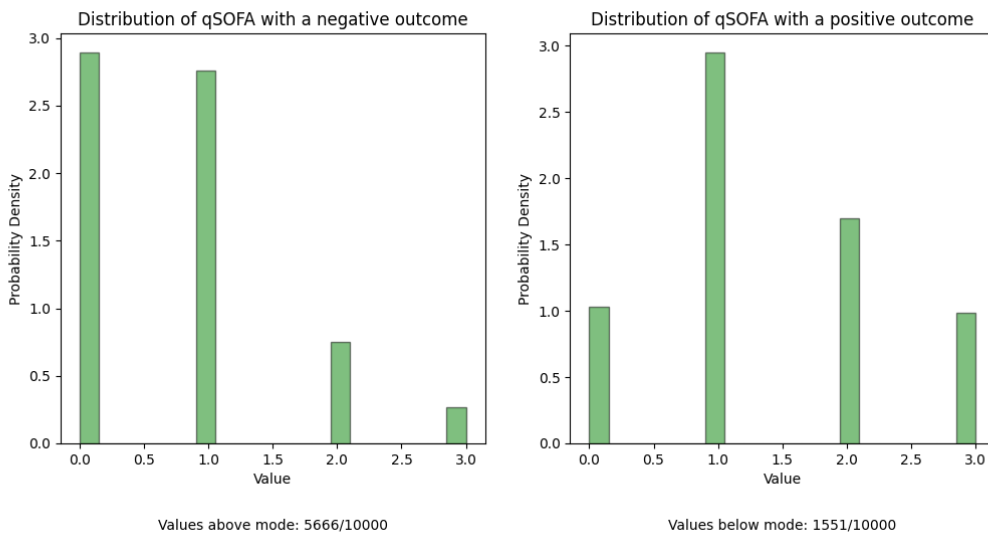

Values above mode: 5666/10000   Values below mode: 1551/10000

Figure 12 – distribution with bounds and standard deviation = 1

Until now it was assumed that the standard deviation is one, but that is probably not true. The standard deviation was found by using knowledge of 2.2.4, as learned from multiple studies the sensitivity of qSOFA was around 69% and the specificity of qSOFA was around 75%. This means that 69% of the positive outcomes/ICU admitted patients have a qSOFA score that is equal to or higher than the mode which is one and therefore 31% below the mode. The same holds for negative outcomes/patients who were not admitted to the ICU. 75% of the qSOFA scores should be lower or equal to the mode which is zero and therefore 25% should be above the mode. The standard deviations, 1.85 for negative outcomes and 1.28 for positive outcomes were found which gave these distributions.
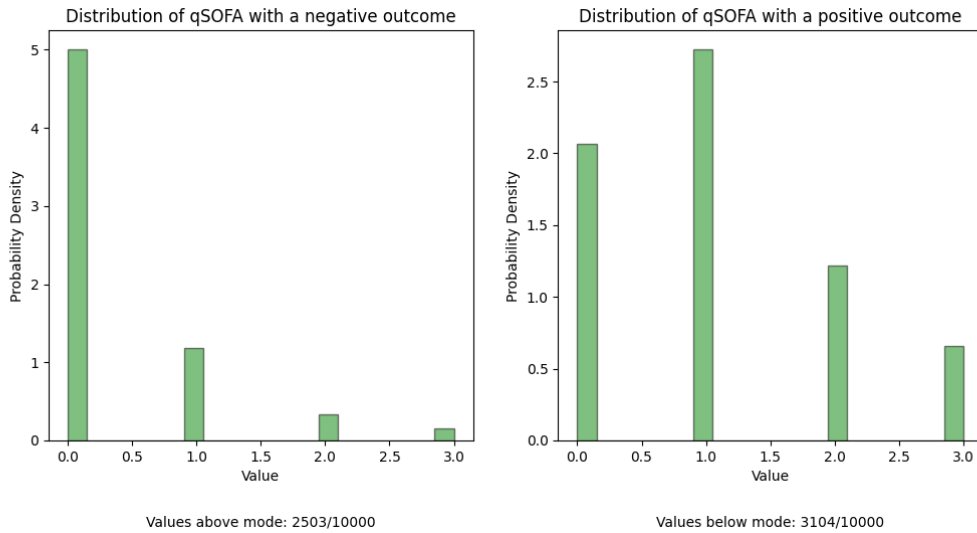
Figure 13 – distribution with bounds and tailored standard deviations

4.4 Data processing

Now that both a qSOFA score simulation and a patient journey dataset have been achieved. The data should be processed and combined. With Python, a program was written to determine the outcome of all traces. If a lab appointment was in the trace, it was given an outcome of one, positive, and if not then it was given a 0, negative. Now all data points were assigned to the traces. All traces with an outcome of one were given a qSOFA score from the positive outcome set and all traces with an outcome of zero were given a qSOFA score from the negative outcome set. When a qSOFA score value was attached to a trace it was erased from the dataset such that it would not appear twice, and all data points were attached to a trace. All qSOFA scores were added to a completed admission event and all outcomes were given to a completed billing event as the billing event is always the last event in a trace. In figure 14, you see the data that was finalized and attached to the event log.
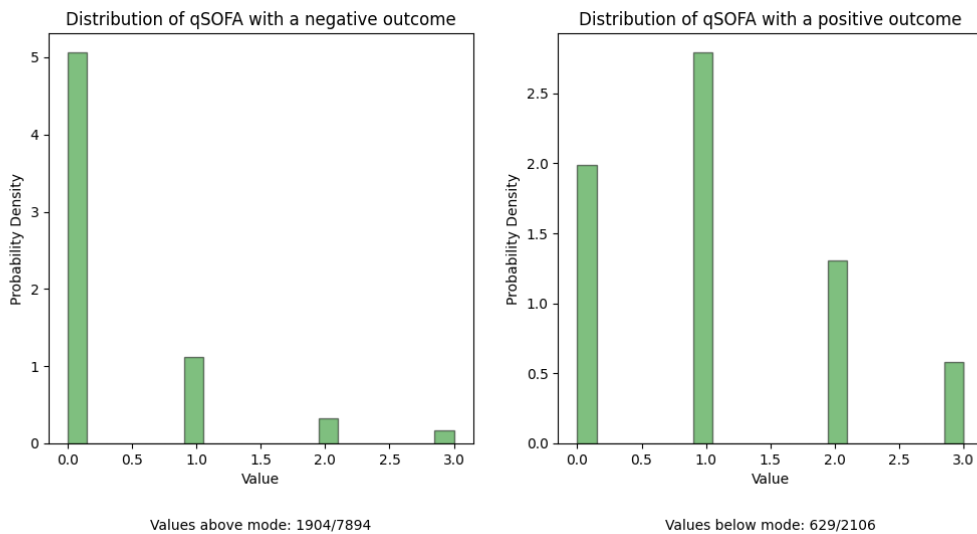

Figure 14 – The utilized final distributions

4.5 Data analysis process

When all data was combined and processed in such a way that it was ready for analysis, Prom Lite 1.4 by Eindhoven University of Technology was used to generate results. The results were generated by the "Perform Predictions of Business Process Features" by Massimiliano de Leoni. In the plugin, the qSOFA score and the outcome were selected which the plugin used to build a decision tree which is the same as in figure 2. In the tool, the two important variables the qSOFA score and the outcome were selected, and an augmented event log was made. Then in the configuration tab dependent variable was set to outcome and some experimentations were done with the visualization mode to get the various results seen in chapter 5.

Summary

In chapter 4 the struggle to obtain real data was briefly explained. This was followed by the process that was used in this project to obtain results. Firstly, a patient flow dataset from general practitioner patients underwent a discovery process such that it could be used to simulate a patient flow through a hospital for which the lab appointment event was chosen to simulate an ICU admission. Then two distributions were made that simulate qSOFA-to-outcome data and were then attached to the patient flow dataset. Furthermore, the mining & analysis process utilizing Prom Lite 1.4 was briefly explained which is how the results were generated.

5. Analysis

In this chapter, the results are shown and analysed using multiple visualisation styles accompanied by a brief explanation of what the results mean and how they should be interpreted. Furthermore, two predictor validation subquestions were formulated which are answered in this chapter:

Subquestion 5: Is the prediction model made with Prom Lite 1.4 accurate?
Subquestion 6: Is the prediction model made with Prom Lite 1.4 scientifically valid?

Firstly, the results are shown in the figures and tables below. After this, the results are discussed subquestions 5 and 6 are answered.
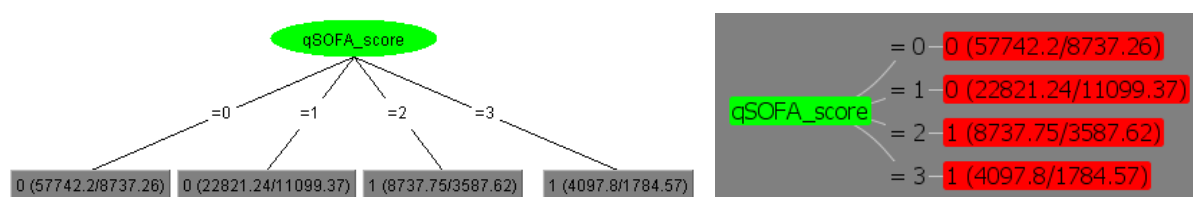

Figure 15 – Decision tree in helicopter- (Left) and explorative- (Right) visualisation

Title	Number	Percentage
Correctly Classified Instances	68943	73.8156 %
Incorrectly Classified Instances	24456	26.1844 %
Kappa statistic	0.2379	-
Mean absolute error	0.2318	-
Root mean squared error	0.3372	-
Relative absolute error	-	84.0478 %
Root relative squared error	-	90.7919 %
Total Number of Instances	93399	-

Table 3 – Data encompassing the generated decision model

A	B	<-- classified as
61756	4343	a = 0
20113	7187	b = 1

Table 4 – Confusion matrix encompassing the generated decision model

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.934	0.737	0.754	0.934	0.835	0.273	0.733	0.834	0
0.263	0.066	0.623	0.263	0.370	0.273	0.733	0.517	1
0.738	0.541	0.716	0.738	0.699	0.273	0.733	0.742	Wgt. Avg.

Table 5 – Detailed accuracy per class encompassing the generated decision model

The model that was generated by Prom Lite 1.4 has an accuracy of 73.8% (from Table 3) which is the answer to subquestion 5. Interestingly the model found the best cutoff at qSOFA ≥ 2 instead of qSOFA ≥ 1 which can be derived from figure 15 as the qSOFA scores 0 & 1 were linked to a negative outcome (0) and the qSOFA scores 2 & 3 are linked to a positive outcome (1). This means that the qSOFA should be equal to 2 or higher (qSOFA ≥ 2) to get a positive outcome which is an ICU admission. This might be because ICU admissions themselves are less common than a person without ICU admission needs with a qSOFA = 1.

Going a bit deeper into the accuracy of the model using the statistics of Table 5 gives us a more detailed view of the accuracy. The fact that the cutoff qSOFA ≥ 2 has been formulated by the algorithm seems to lead to a very high FP (False Positive) rate, 0.737, in Class 0. Many instances that were of a positive outcome (outcome = 1) were put into the 0 class. Similarly, the TP (true positive) rate for Class 1 suffers as the value is 0.263. Concluding this model is probably too insensitive for positive outcomes given that the simulated data is a good representative of the actual distributions.

The F-measure which combines precision and recall, also tells us a bit more about the performance. Again, the performance for positive cases, Class 1, suffers as this measure is 0.370. Using Matthews Correlation Coefficient (MCC) performance is measured against randomness = 0 and here the model is better than random, but with just 0.273, low. Using the ROC and PRC area performance values gives the same conclusion as MCC there is something as both values are significantly above 0.5, but 0.733 and 0.742 are both quite far from the ideal which is 1. Yes, the model is quite accurate at 73.8%, but taking performance measures into account the reliability and reproducibility of this high accuracy is not as strong as hoped.

Taking some more performance statistics from Table 3, the already established image of a potentially unreliable performance is confirmed further. The inter-rater reliability measured by the kappa statistic is 0.2379 concludes as poor using the guidelines of Fleiss (1981). Root relative squared error (RRSE) 90.8%, also gives a significant indication of poor performance. Both the kappa statistic and RRSE are red flags that the simulated data might not be an accurate representation of the real qSOFA score data. To answer subquestion 6, the ICU admission predictor generated by the algorithm cannot be taken as scientifically valid as poor performance has been flagged by a multitude of performance measurements.

Summary

In this chapter, the validation subquestions 5 and 6 were formulated and answered. It was found that the algorithm of the "Perform Predictions of Business Process Features" plugin chose qSOFA ≥ 2 as the most accurate ICU admission predictor which was different from the qSOFA ≥ 1 obtained from background literature. The accuracy of qSOFA ≥ 2 was 73.8% which is close to what qSOFA ≥ 1 was in the literature. After taking multiple performance metrics into account, it was concluded that the ICU admission predictor qSOFA ≥ 2 performs too poorly to be taken as scientifically valid given that the simulated data is an accurate representation of the real data.

6. Discussion

As most research projects, this project has its fair share of limitations and imperfections. The largest limitation that led to a multitude of imperfections and points of improvement was the fact that the MST hospital was not able to deliver real data for us to analyse. This resulted in the creation of a simulated and combined dataset that lacked depth and created potential time-wise inaccuracy and potential inaccuracy in the qSOFA-to-outcome data.

Firstly, the self-created dataset lacked depth. This means that there were no secondary factors to analyse to get a deeper understanding of the patients. An example would be that patients with different blood types might have different chances of landing in the ICU or that patients who get a qSOFA spike during phase X of the process are more likely to not end up in the ICU or that qSOFA-scores can only accurately predict ICU admissions in a certain sex. These are just examples to show that with a full and more elaborated dataset, it would have been possible to look way deeper into deeper underlying correlations which is exactly where process mining technologies seem to shine.

There is another factor in the self-created dataset that needs discussion. The lab appointment event was taken as a placeholder for an ICU admission. In the dataset, there were a total of 2106 events with at least one lab appointment out of 10000. Whether the 21.1% ICU admission rate is accurate is highly debatable. According to Barrett et al. (2014), the rate in 2011 in the USA was 26.9%. However, in contrast, when dividing the amount of Dutch ICU admissions of Stichting NICE (2023) by the amount of Dutch hospitalisation by the Centraal Bureau voor Statistiek (2023), the ICU admission rate was only 2.1% in 2013 and more recently, in 2020, the rate was 2.5%. Due to highly differentiating healthcare systems, this rate differs significantly per country. However, this rate does influence the results of this research project. The smaller the size of the ICU admission, and positive outcome pool the stronger the influence of the noise from the negative outcome pool as the chance that high qSOFA score values still come from the negative outcome pool becomes higher. Also, with a smaller positive outcome pool the chance of and influence of potential noise rises.

Time-wise inaccuracy and lack of depth are also a limiting factor in the study. With the simulated dataset it was only possible to insert a single frame of qSOFA score measurement. In a real dataset, the qSOFA score could have been measured in every single step of the process. There is a fair chance that the qSOFA score tells us more about the 48-hour chance on ICU Admission, but not a lot about the 7-day chance on ICU admission or vice versa. Covino et al., (2020), for example, found no significant difference between 7 days and 48 hours for ICU admissions, but a significant difference in predicting ICU mortality between 48 hours and 7 days. The article shows that the time perspective can have a great influence on whether a prediction can be accurate or not.

The qSOFA-to-outcome data also have a significant chance of not being accurately representative of real data. In an effort to smartly simulate qSOFA-to-outcome, all choices have been made on background research combined with educated guessing. This was mainly caused by the fact there exists no open-source data which could help finding how real distributions would look like in a real dataset. This is caused by the fact that every biological and biochemical symptom has vastly different ways of developing. This is clearly represented in scientific literature as every disease or medication needs its own study.

Another limitation worth mentioning is the qSOFA score itself. The qSOFA-score has shown some results to be a predictor of ICU admissions, but none of the studies have shown accuracy

rates above 90% for any predictor of ICU admissions. Furthermore, the qSOFA-score might be too binary. Meaning that of the three inputs, there is one binary outcome that gets added up. If the qSOFA-score would be more scalar, it might be significantly more efficient in predicting ICU admissions.

7. Concluding remarks

In this project, utilizing process mining technologies for predicting ICU Admissions using the qSOFA score was investigated, as in the Netherlands the costs of ICUs were estimated to be over a billion dollars. Predicting and spotting potential ICU admissions could reduce costs in terms of money as well as resources. Furthermore, early predictions could also lead to improved clinical outcomes for patients and reduce their mortality rates when progressions of medical conditions can be prevented.

Utilizing the strategy proposed by Kitchenham et al. (2009) subquestion 1: "What are the current best ways to predict ICU admissions?" was answered. Six papers were selected (Mahammadova et al., (2023); Cresti et al., (2023); Feghoul et al., (2023); Naser et al., (2023); Zahran et al. (2023)) and they found different indicators for future ICU admissions. The six papers illustrate that the topic is highly controversial in scientific literature and scientists do not agree on which factors/metrics can be utilized to predict future ICU admissions. Subquestion 2: "How are ICU admissions predicted in practice?" was answered as a reaction to the literature of subquestion 1 combined with common hospital knowledge. To answer subquestion 2, in practice, doctors are using their clinical expertise to decide whether a patient has ICU needs. The reason for this is mainly that science is not sure yet about what are the best ways to predict ICU admissions. Also, there is an ethical reason for this decision. Right now, society prefers that a human is responsible for making critical decisions for patients. Subquestion 3: "How is the qSOFA score related to ICU admissions?" is also answered by utilizing the strategy proposed by Kitchenham et al. (2009). Answering subquestion 3, the qSOFA score is a highly promising measurement for predicting ICU admissions as its accuracy for qSOFA ≥ 1 is determined by the average of the sensitivity and specificity over four separate studies (Zhang et al., 2020; Bae et al., 2022; Chu et al. (2020); Covino et al. (2020)) is 72%. Even though, metrics like the qSOFA score exist, it was found that those metrics are not used in practice. Presumably, this is caused by the fact that the accuracy of all methods is still too low to provide consistent value in hospitals. The last background research question: "How can process mining be used for making ICU admission predictions?" was mainly answered with knowledge obtained from the book written by Van der Aalst (2016). It became clear that process mining is an upcoming set of techniques in science with high potential in scientific fields where a process has a significantly consistent structure such that analysis over time is valuable. Utilizing the prediction strategies proposed by Van der Aalst (2016), predictions can be made using process mining technologies which is the answer to subquestion 4.

The PM^2 method was chosen to give structure to this research project and phases 1 to 5 were executed in the project. Phase 6 "Process Improvement and Support" of the PM^2 method was the only phase that was outside the scope of this project. However, recommendations for future work are made in the discussion which can be seen as a part of a version of phase 6.

Getting data was the part where this research project did not go as planned as the MST Hospital was not able to deliver data in time. After trying to get real data from many sources for some weeks, an alternative approach was formulated which utilized a patient flow dataset of general practitioner patients combined with simulated qSOFA-to-outcome data. The data was then analysed using the plugin "Perform Predictions of Business Process Features" from Prom Lite 1.4. The output of this plugin was used to answer the validation questions which are subquestions 5 and 6 which can be found in figure 15, table 3, table 4 and table 5.

The resulting ICU admission predictor was qSOFA ≥ 2 which was different from qSOFA ≥ 1 which scientific literature gave us. Answering subquestion 5, the accuracy of the predictor,

qSOFA ≥ 2, was 73.8% which is close to the accuracy of qSOFA ≥ 1 according to scientific literature which was 72%. When looking at a multitude of performance metrics from the results it must be concluded that the performance of the model is too poor to take the prediction model as scientifically valid which is the answer to subquestion 6.

In chapter 6, the multiple limitations were discussed. Firstly, the self-created dataset lacked depth. Secondly, taking the lab appointment as a placeholder for the ICU admission might have influenced the results as it is unsure whether the occurrence rate of 2106 out of 10000 is realistic. Time-wise inaccuracy and lack of depth due to not having a real data set were also noted as limitations. In chapter 6, the conclusion that the qSOFA-to-outcome data might not be representative of the real data was also made. The last limitation that was noted is the qSOFA score itself as no studies found accuracies over 90% and the qSOFA score might be too binary which might be a reason that high accuracies cannot be reached yet.

To answer the main research question, yes, an ICU admission predictor can be generated using process mining techniques and qSOFA data. Even though the process mining field has high potential, and the accuracy of the predictor was quite high, the strategy carried out in this paper did not produce results that can be counted as scientifically valid.

For future work, it would be most valuable to obtain an actual dataset as it can fully unlock the potential of process mining technologies. Also, answering the question of what the rate of ICU admissions to non-ICU admissions should be, would be helpful in future work. Furthermore, redesigning the qSOFA-score such that the score becomes more scalar rather than binary might result in more accurate predictions. Another potential recommendation for future work would be to take a totally different metric as a predictor for ICU admissions. Van Mourik et al. (2023) found that taking a MEWS ≥ 6 (Modified Early Warning Score) led to an 82.5% accuracy in predicting ICU admissions in their paper. Many more predictive measurements have been formulated in scientific literature. Combining those with process mining has mostly not been done and might yield strong results.

References

Bae, S. J., Lee, J. H., & Choi, Y. H. (2022). Comparison of Predictive Scoring Systems in Assessing Risk for Intensive Care Unit Admission and In-Hospital Mortality in Patients with Urinary Tract Infections. Journal of Critical and Intensive Care. https://doi.org/10.37678/dcybd.2022.2941

Barrett, M. L., Smith, M. W., Elixhauser, A., Honigman, & Pines, J. M. (2014). Utilization of Intensive Care Services, 2011. In Healthcare Cost and Utilization Project. Retrieved March 22, 2024, from http://www.hcup-us.ahrq.gov/reports/statbriefs/sb185-Hospital-Intensive-Care-Units-2011.pdf

Bentrem, D. J., Yeh, J. J., Brennan, M. F., Kiran, R. P., Pastores, S. M., Halpern, N. A., Jaques, D. P., & Fong, Y. (2005). Predictors of intensive care unit admission and related outcome for patients after pancreaticoduodenectomy. Journal of Gastrointestinal Surgery, 9(9), 1307–1312. https://doi.org/10.1016/j.gassur.2005.09.010

Centraal Bureau voor Statistiek. (2023). Ziekenhuisopnamen en -patiënten; diagnose-indeling VTV [Dataset]. https://opendata.cbs.nl/statline/#/CBS/nl/dataset/84067NED/table?dl=95AEE

Chu, S., Seak, C., Su, T., Chaou, C., Tseng, H., & Li, C. (2020). Prognostic accuracy of SIRS criteria and qSOFA score for in-hospital mortality among influenza patients in the emergency department. BMC Infectious Diseases, 20(1). https://doi.org/10.1186/s12879-020-05102-7

Chioncel, O., Bubenek, Ş., Ambrosy, A. P., Macarie, C., & Gheorghiade, M. (2012). Clinical Predictors of Intensive Care Unit Admission in Patients Hospitalized for Pulmonary Edema–Data from the RO-AHFS Registry. Journal of Cardiac Failure (Print), 18(8), S99. https://doi.org/10.1016/j.cardfail.2012.06.380

Covino, M., Sandroni, C., Santoro, M., Sabia, L., Simeoni, B., Bocci, M. G., Ojetti, V., Candelli, M., Antonelli, M., Gasbarrini, A., & Franceschi, F. (2020). Predicting intensive care unit admission and death for COVID-19 patients in the emergency department using early warning scores. Resuscitation, 156, 84–91. https://doi.org/10.1016/j.resuscitation.2020.08.124

Cresti, A. (2023). Clinical Features and Mortality Rate of Infective endocarditis in Intensive Care Unit: A Large-Scale Study and Literature Review. The Anatolian Journal of Cardiology, 44–54. https://doi.org/10.14744/anatoljcardiol.2023.3463

Dasta, J. F., McLaughlin, T., Mody, S. H., & Piech, C. T. (2005). Daily cost of an intensive care unit day: The contribution of mechanical ventilation*. Critical Care Medicine, 33(6), 1266–1271. https://doi.org/10.1097/01.ccm.0000164543.14619.00

Data in beeld. (2023). stiching-nice.nl. Retrieved November 16, 2023, from https://www.stichting-nice.nl/datainbeeld/public?subject=BASIC&year=2022&hospital=-1&icno=0

DeMers, D. (2023, April 10). Physiology, mean arterial pressure. StatPearls - NCBI Bookshelf. https://www.ncbi.nlm.nih.gov/books/NBK538226/

Dreamshaper 8. (2024, February 25). Prediciting ICU admissions. dezgo.com. https://dezgo.com/txt2img

Feghoul, L., Caillault, A., Peyrony, O., Salmona, M., Néré, M., Delaugerre, C., Azoulay, É., Chevret, S., & Goff, J. L. (2023). Respiratory torque teno virus load at emergency department visit predicts intensive care unit admission of SARS‑CoV‑2 infected patients. Journal of Medical Virology, 95(12). https://doi.org/10.1002/jmv.29319

Fleiss, J. L. (1981). Statistical Methods for Rates and Proportions, 2nd ed. American Journal of Psychiatry, 138(12), 1644–1645. https://doi.org/10.1176/ajp.138.12.1644-a

Verbeek, H.M.W. & Buijs, J. (n.d.), Introduction to Process Mining with ProM [MOOC]. FutureLearn. https://www.futurelearn.com/courses/process-mining

Harmon. (2010). Business Process Change (second edition) [Pdf]. Elsevier. https://books.google.nl/books?hl=en&lr=&id=wSLnCTBjbn8C&oi=fnd&pg=PP1&ots=20K-VLGpdA&sig=T1DZitOeA0FAuhcZRvnOj2Aaf_M&redir_esc=y#v=onepage&q&f=false

Holguín, F., Ramadan, B., Gal, A. A., & Roman, J. (2008). Prognostic factors for hospital mortality and ICU admission in patients with ANCA-Related Pulmonary Vasculitis. The American Journal of the Medical Sciences, 336(4), 321–326. https://doi.org/10.1097/maj.0b013e31816805fa

Ju, J., Zhang, P., Wang, Y., Kou, Y., Fu, Z., Jiang, B., & Zhang, D. (2021). A clinical nomogram predicting unplanned intensive care unit admission after hip fracture surgery. Surgery, 170(1), 291–297. https://doi.org/10.1016/j.surg.2021.01.009

Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering – A systematic literature review. Information and Software Technology, 51(1), 7–15. https://doi.org/10.1016/j.infsof.2008.09.009

Mahammadova, S., Oba, S., Çinar, A. S., Türk, H. Ş., & Özdemi, H. M. (2023). Perioperative characteristics and predictors of intensive care needs in patients undergoing surgery for scoliosis: a retrospective study of 165 patients. Signa Vitae, dafa2394-3198-48c3-ba35-56c27b26c71440db3e36-1990-43e3-923a-9f364549568d. https://doi.org/10.22514/sv.2023.121

Monaco, F., Di Tomasso, N., Landoni, G., Nardelli, P., Radinovic, A., Melillo, F., D'Angelo, G., Della Bella, P., Zangrillo, A., & Mazzone, P. (2019). Predictors of intensive care unit admission in patients undergoing lead extraction: A 10-Year observational study in a High-Volume center. Journal of Cardiothoracic and Vascular Anesthesia (Print), 33(7), 1845–1851. https://doi.org/10.1053/j.jvca.2019.02.012

Moreno, R. P., & Metnitz, P. (2008). Severity Scoring Systems: tools for the evaluation of patients and intensive care units. In Elsevier eBooks (pp. 1547–1565). https://doi.org/10.1016/b978-032304841-5.50076-5

Naser, A. Y., Dairi, M. S., Alwafi, H., Ashoor, D. S., Qadus, S., Aldhahir, A. M., Alqarni, A., Badr, O. I., Qanash, S., Hafiz, W., Alqahtani, J. S., Ekram, R., Abuirmeileh, A., Jarab, A. S., & Badr, O. I. (2023). The rate of ward to intensive care transfer and its predictors among hospitalized COPD patients, a retrospective study in a local tertiary center in Saudi Arabia. BMC Pulmonary Medicine, 23(1). https://doi.org/10.1186/s12890-023-02775-z

Oduncu, A. F., Kıyan, G. S., & Yalçınlı, S. (2021). Comparison of qSOFA, SIRS, and NEWS scoring systems for diagnosis, mortality, and morbidity of sepsis in emergency department. American Journal of Emergency Medicine, 48, 54–59. https://doi.org/10.1016/j.ajem.2021.04.006

Qin, N., Wang, J., Wu, X., Wang, W., Wang, X., Peng, X., & Wang, L. (2023). Characterization and prognosis of autoimmune encephalitis in the neurological intensive care unit: a retrospective study. Neurological Sciences, 44(8), 2889–2895. https://doi.org/10.1007/s10072-023-06778-w

Van Eck, M. M., Lu, X., Leemans, S. J. J., & Van der Aalst, W. M. P. (2015). PM2: A Process Mining Project Methodology. In Lecture Notes in Computer Science (pp. 297–313). https://doi.org/10.1007/978-3-319-19069-3_19

Van Mourik, N., Oomen, J. J., Van Vught, L. A., Biemond, B. J., Van Den Bergh, W. M., Blijlevens, N. M. A., Vlaar, A. P., & Müller, M. C. (2023). The predictive value of the modified early warning score for admission to the intensive care unit in patients with a hematologic malignancy – A multicenter observational study. Intensive and Critical Care Nursing, 79, 103486. https://doi.org/10.1016/j.iccn.2023.103486

Van der Aalst, W. (2016). Process mining : data science in action (Second edition). Springer. https://doi.org/10.1007/978-3-662-49851-4

Vincent, J., Moreno, R. P., Takala, J., Willatts, S. M., De Mendonça, A., Bruining, H. A., Reinhart, C. K., Pm, S., & Thijs, L. G. (1996). The SOFA (Sepsis-related Organ Failure Assessment) score to describe organ dysfunction/failure. Intensive Care Medicine, 22(7), 707–710. https://doi.org/10.1007/bf01709751

Society of Critical Care Medicine. (n.d.). SCCM | Critical Care Statistics. Society of Critical Care Medicine (SCCM). Retrieved April 3, 2024, from https://sccm.org/Communications/Critical-Care-Statistics

Seymour, C. W., Vincent, J., Iwashyna, T. J., Brunkhorst, F. M., Rea, T. D., Scherag, A., Rubenfeld, G. D., Kahn, J. M., Shankar-Hari, M., Singer, M., Deutschman, C. S., Escobar, G. J., & Angus, D. C. (2016). Assessment of clinical criteria for sepsis. JAMA, 315(8), 762. https://doi.org/10.1001/jama.2016.0288

Zahran, T. E., Kalot, N., Cheaito, R., Khalifeh, M., Estelly, N., & Majzoub, I. E. (2023). Predictors of intensive care unit admission in adult cancer patients presenting to the emergency department with COVID-19 infection: A retrospective study. PLOS ONE, 18(8), e0287649. https://doi.org/10.1371/journal.pone.0287649

Zhang, X., Liu, B., Liu, Y., Ma, L., & Hong, Z. (2020). Efficacy of the quick sequential organ failure assessment for predicting clinical outcomes among community-acquired pneumonia patients presenting in the emergency department. BMC Infectious Diseases, 20(1). https://doi.org/10.1186/s12879-020-05044-0

4TU.ResearchData. (2023). Healthcare treatment process (logs and CPN model)- Data underlying the publication" Identifying the context of data usage to diagnose privacy issues through Process Mining" (Version V1) [Dataset]. https://doi.org/10.4121/uuid:d9769f3d-0ab0-4fb8-803b-0d1120ffcf54

Appendices

Appendix 1: Code of the event log editor

```python
import xml.etree.ElementTree as ET
import csv

# working with xml files
tree = ET.parse('processLog_updated.xes')
root = tree.getroot()

# Load data from csv
zero_array = []
one_array = []
file = open("qSOFAtoOutcomeData_final.csv", "r")
csv_read = csv.reader(file, delimiter=",")
dataset = list(csv_read)
file.close()

# Counters
amount_of_zeros = 0
amount_of_ones = 0
amount_of_neg_outcomes = 0
amount_of_pos_outcomes = 0
amount_of_traces = 0
amount_of_billings = 0
lab_test_is_true_count = 0
filled_zero_array_length = 0
filled_one_array_length = 0


# Finds the outcome of a current trace
def find_outcome(current_trace):
    for current_trace_string in current_trace.findall(f'string'):
        current_trace_key = current_trace_string.get('key')
        current_trace_value = current_trace_string.get('value')
        if current_trace_key == "Outcome" and current_trace_value == "0":
            return 0
        if current_trace_key == "Outcome" and current_trace_value == "1":
            return 1


# a trace counter for debugging
for trace in root.findall('trace'):
    amount_of_traces += 1

# a billing event counter for debugging
for trace in root.findall('trace'):
    for event in trace.findall('event'):
        previous_string = ''
        for string in event.findall('string'):
            key = string.get('key')
            value = string.get('value')
            parent = string.get
            if key == 'lifecycle:transition' and value == 'COMPLETE' and previous_string.get(
                    'key') == 'concept:name' and
previous_string.get('value') == 'Billing (bi)':
                amount_of_billings += 1
            previous_string = string
```

```python
for value_set in dataset:
    if value_set[1] == '0':
        zero_array.append(value_set[0])
    if value_set[1] == '1':
        one_array.append(value_set[0])

filled_zero_array_length = len(zero_array)
filled_one_array_length = len(one_array)

# This loop adds a outcome variable to the xml file if there is a lab
appointment in the trace it adds a 1 if not
# then it adds a 0
for trace in root.findall('trace'):
    lab_test = False
    for event in trace.findall('event'):
        previous_string = ''  # initiate a previous string such that it
saves the previous string
        for string in event.findall('string'):
            key = string.get('key')
            value = string.get('value')
            if key == 'lifecycle:transition' and value == 'COMPLETE' and
previous_string.get(
                    'key') == 'concept:name' and
previous_string.get('value') == 'Lab appointment (la)':
                print('la detected!')
                lab_test = True
                lab_test_is_true_count += 1
            if key == 'lifecycle:transition' and value == 'COMPLETE' and
previous_string.get(
                    'key') == 'concept:name' and
previous_string.get('value') == 'Billing (bi)':
                complete_billing = event
            previous_string = string
    if lab_test == False:
        ET.SubElement(trace, 'string', attrib={'key': 'Outcome', 'value':
'0'})
        ET.SubElement(complete_billing, 'string', attrib={'key': 'Outcome',
'value': '0'})
        amount_of_neg_outcomes += 1
        print('Outcome: Negative')
    elif lab_test == True:
        ET.SubElement(trace, 'string', attrib={'key': 'Outcome', 'value':
'1'})
        ET.SubElement(complete_billing, 'string', attrib={'key': 'Outcome',
'value': '1'})
        amount_of_pos_outcomes += 1
        print('Outcome: Positive')

# attaches qSOFA-scores from the array of from the simulated qSOFA dataset
for trace in root.findall('trace'):
    for event in trace.findall('event'):
        for string in event.findall('string'):
            key = string.get('key')
            value = string.get('value')
            if key == 'lifecycle:transition' and value == 'COMPLETE':
                if previous_string.get('key') == 'concept:name' and
previous_string.get('value') == 'Admission (ad)':
                    outcome = find_outcome(trace)
                    if outcome == 0:
                        ET.SubElement(event, 'string', attrib={'key':
'qSOFA-score', 'value': str(zero_array[0])})
```

```python
                            amount_of_zeros += 1
                            zero_array.pop(0)
                        if outcome == 1:
                            ET.SubElement(event, 'string', attrib={'key':
'qSOFA-score', 'value': str(one_array[0])})
                            one_array.pop(0)
            previous_string = string

# copy existing traces with outcome = 0 such that pos outcome is 2.561%
trace_count = 10000
while trace_count < 82234:
    if trace_count == 82234:
        break
    for trace in root.findall('trace'):
        if trace_count == 82234:
            break
        for event in trace.findall('event'):
            if trace_count == 82234:
                break
            for string in event.findall('string'):
                if trace_count == 82234:
                    break
                key = string.get('key')
                value = string.get('value')
                if key == 'Outcome' and value == '0':
                    ET.SubElement(root, 'trace', trace)
                    trace_count += 1



tree.write('processLog_updated_finalV2_2561.xml')

print("Trace count: " + str(amount_of_traces))
print("Billing count: " + str(amount_of_billings))
print("Lab_test is true count: " + str(lab_test_is_true_count))
print("Pos outcome count: " + str(amount_of_pos_outcomes), "&", "Neg
outcome count: " + str(amount_of_neg_outcomes))
print("Filled zero_array length: " + str(filled_zero_array_length), "&",
      "Filled one_array length: " + str(filled_one_array_length))
print("Emptied zero_array length: " + str(len(zero_array)), "&", "Emptied
one_array length: " + str(len(one_array)))
print("trace count: " + str(trace_count))


# Finish
def print_hi(name):
    # Use a breakpoint in the code line below to debug your script.
    print(f'Hi {name} the program has finished!')  # Press Ctrl+F8 to
toggle the breakpoint.


# Press the green button in the gutter to run the script.
if __name__ == '__main__':
    print_hi('Jelle')
```

Appendix 2: Code of the distribution maker

```python
import numpy as np
import matplotlib.pyplot as plt
import csv
from scipy.stats import lognorm

# Distribution for negative outcomes
mode1 = 0
sigma1 = 1.9 #1.85
max_value1 = 3
sample_size1 = 10000

# Distribution for positive outcomes
mode2 = 1
sigma2 = 1.3 #1.28
max_value2 = 3
sample_size2 = 10000

# Calculate scale parameters to achieve the desired modes
scale1 = np.exp(mode1 - 0.5 * sigma1**2)
scale2 = np.exp(mode2 - 0.5 * sigma2**2)

# Create the log-normal distributions
distribution1 = lognorm(sigma1, scale=scale1)
distribution2 = lognorm(sigma2, scale=scale2)

# Generate random samples
samples1 = np.zeros(sample_size1, dtype=int)
samples2 = np.zeros(sample_size2, dtype=int)
values_above_mode1 = 0
values_below_mode2 = 0

# Negative outcome distribution loop
for i in range(sample_size1):
    sample = round(distribution1.rvs())
    while sample > max_value1 or sample < 0:
        sample = round(distribution1.rvs())
    samples1[i] = sample
    if sample > mode1: #Counts the values above the mode: qSOFA ≥ 1 for
Negative outcomes
        values_above_mode1 += 1

# Positive outcome distribution loop
for i in range(sample_size2):
    sample = round(distribution2.rvs())
    while sample > max_value2 or sample < 0:
        sample = round(distribution2.rvs())
    samples2[i] = sample
    if sample < mode2: #Counts the values above the mode: qSOFA < 1 for
positive outcomes
        values_below_mode2 += 1

# Combine the samples and labels
all_samples = np.concatenate((samples1, samples2))
labels = np.concatenate((np.zeros(sample_size1, dtype=int),
np.ones(sample_size2, dtype=int)))

# Write to CSV file
csv_data = list(zip(all_samples, labels))
csv_columns = ['Sample', 'Label']
```

```python
with open('qSOFAtoOutcomeData_Final.csv', 'w', newline='') as csv_file:
    csv_writer = csv.writer(csv_file)
    csv_writer.writerow(csv_columns)
    csv_writer.writerows(csv_data)

# Plotting the histograms
fig, axs = plt.subplots(1, 2, figsize=(12, 6))

# Negative outcome histogram
axs[0].hist(samples1, bins=20, density=True, alpha=0.5, color='g',
edgecolor='black')
axs[0].set_title('Distribution of qSOFA with a negative outcome')
axs[0].set_xlabel('Value')
axs[0].set_ylabel('Probability Density')

# Positive outcome histogram
axs[1].hist(samples2, bins=20, density=True, alpha=0.5, color='g',
edgecolor='black')
axs[1].set_title('Distribution of qSOFA with a positive outcome')
axs[1].set_xlabel('Value')
axs[1].set_ylabel('Probability Density')

# Print values above the mode for negative outcome distribution
axs[0].annotate(f'Values above mode: {values_above_mode1}/{sample_size1}',
                xy=(0.5, -0.2), xycoords='axes fraction',
                ha='center', va='center', fontsize=10)

# Print values below the mode for positive outcome distribution
axs[1].annotate(f'Values below mode: {values_below_mode2}/{sample_size2}',
                xy=(0.5, -0.2), xycoords='axes fraction',
                ha='center', va='center', fontsize=10)

# Adjust layout to make room for the bottom text
plt.subplots_adjust(bottom=0.2)

plt.show()
```

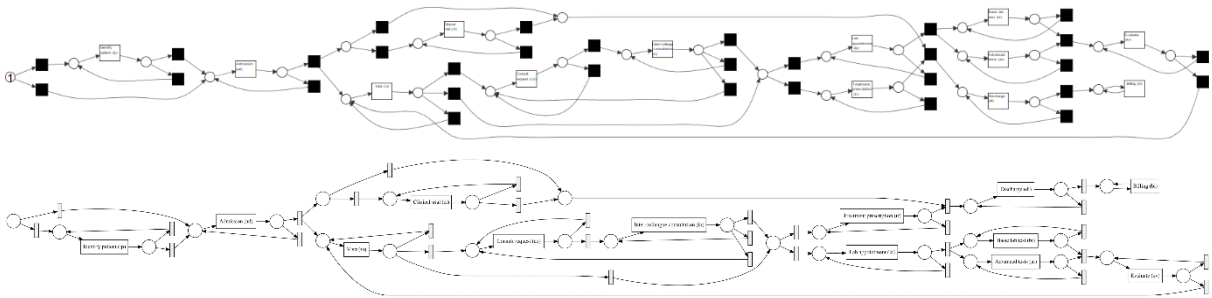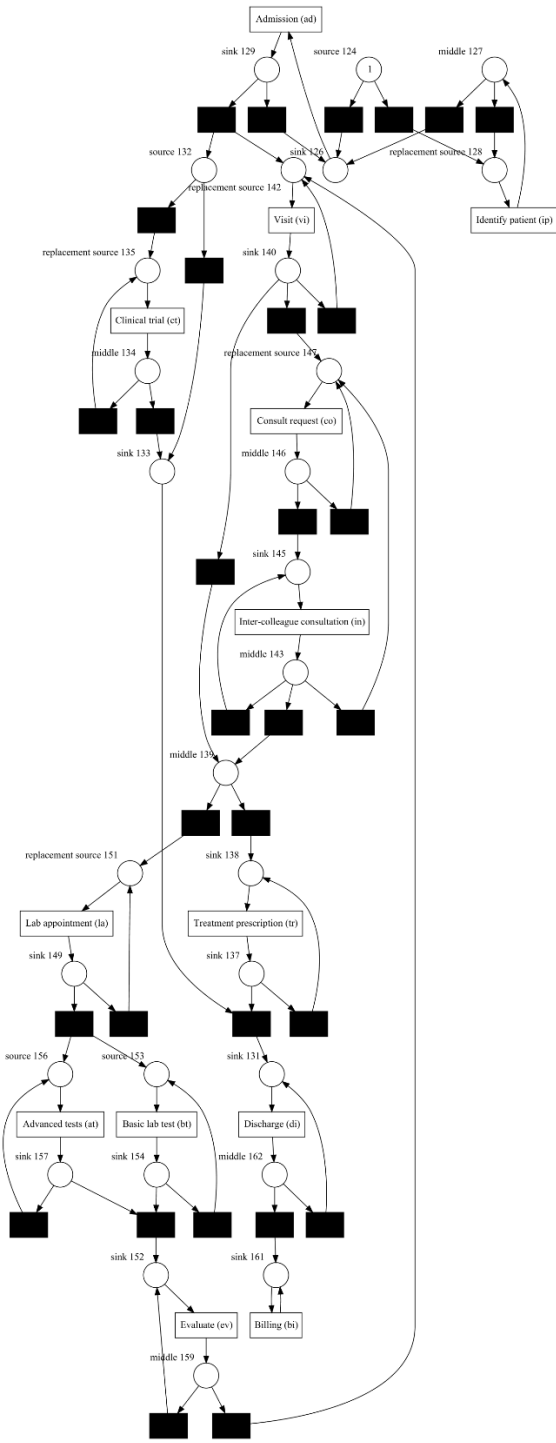# Appendix 3: More process models with Prom Lite 1.4
Alpha miner
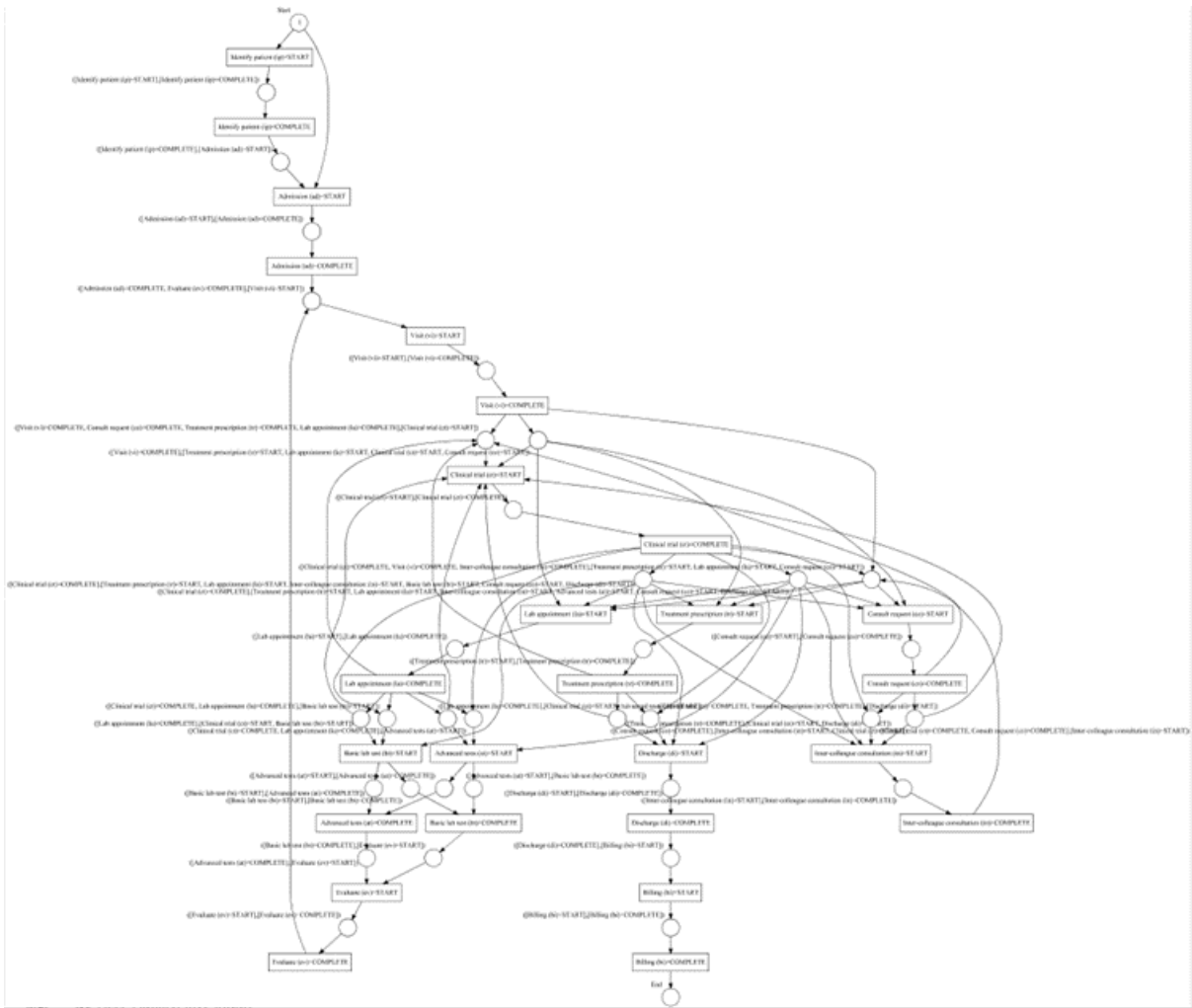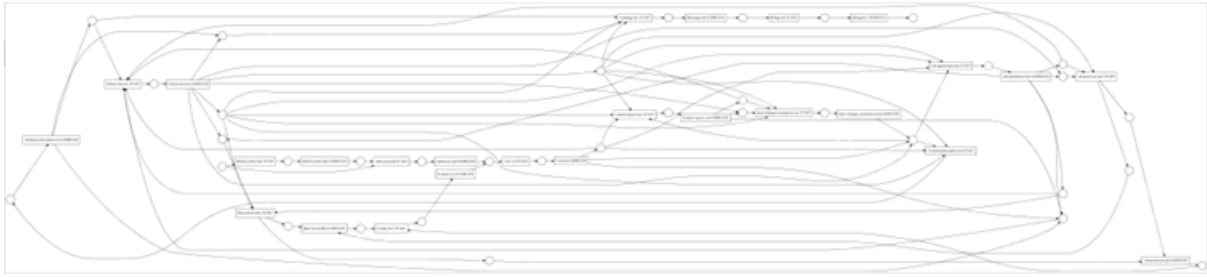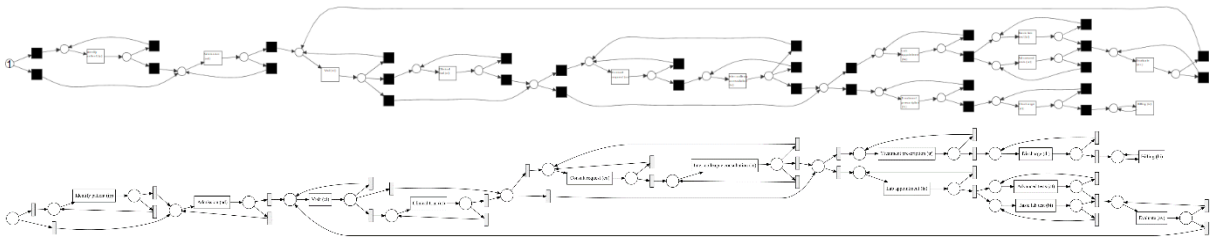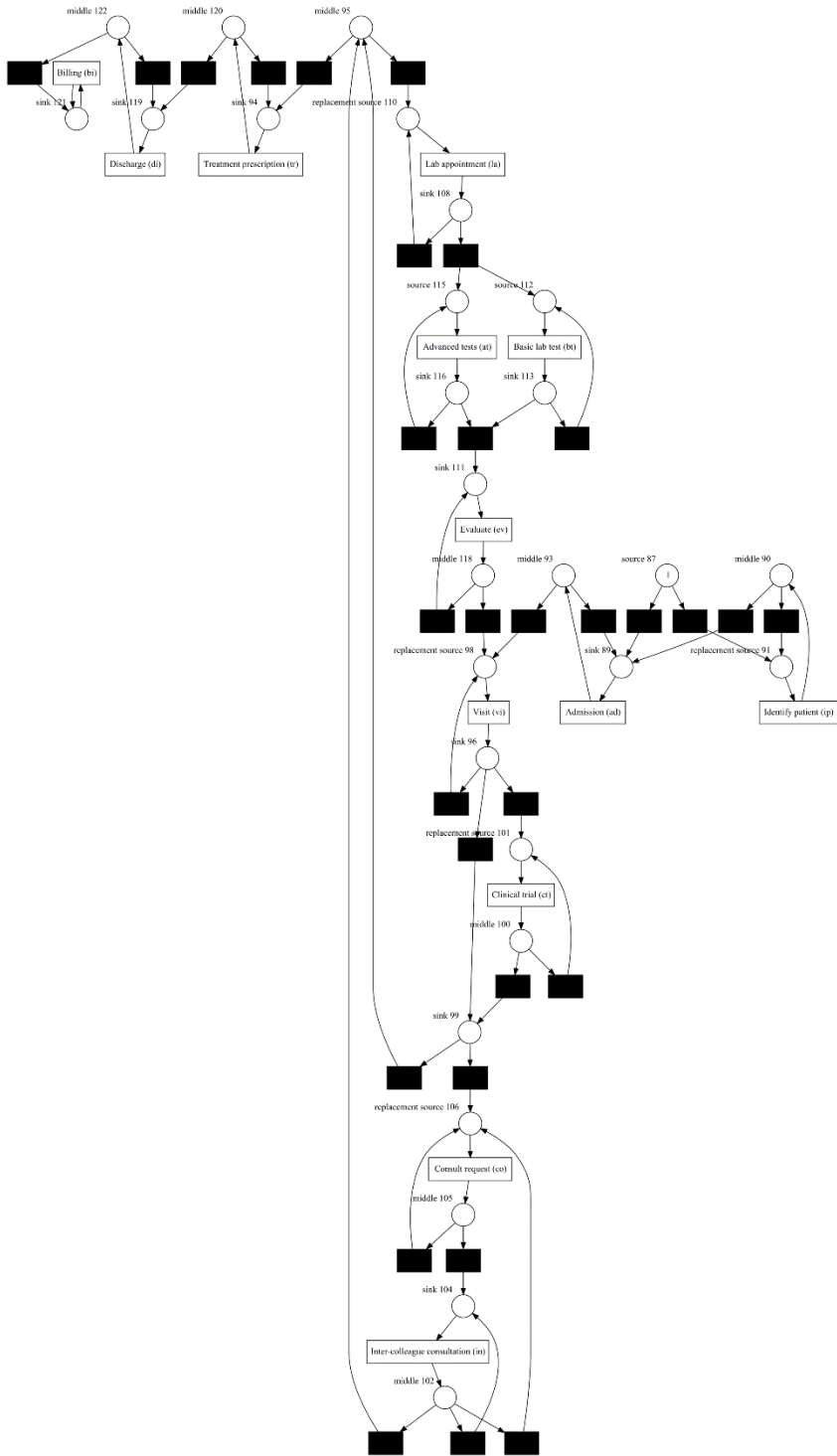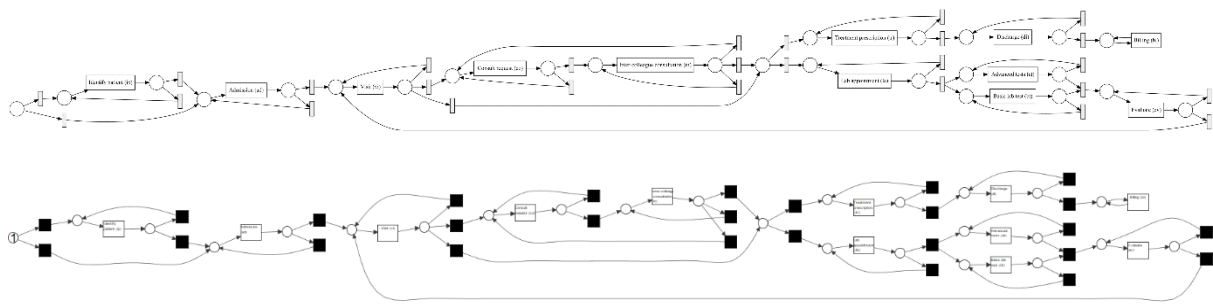
Alpha+ miner
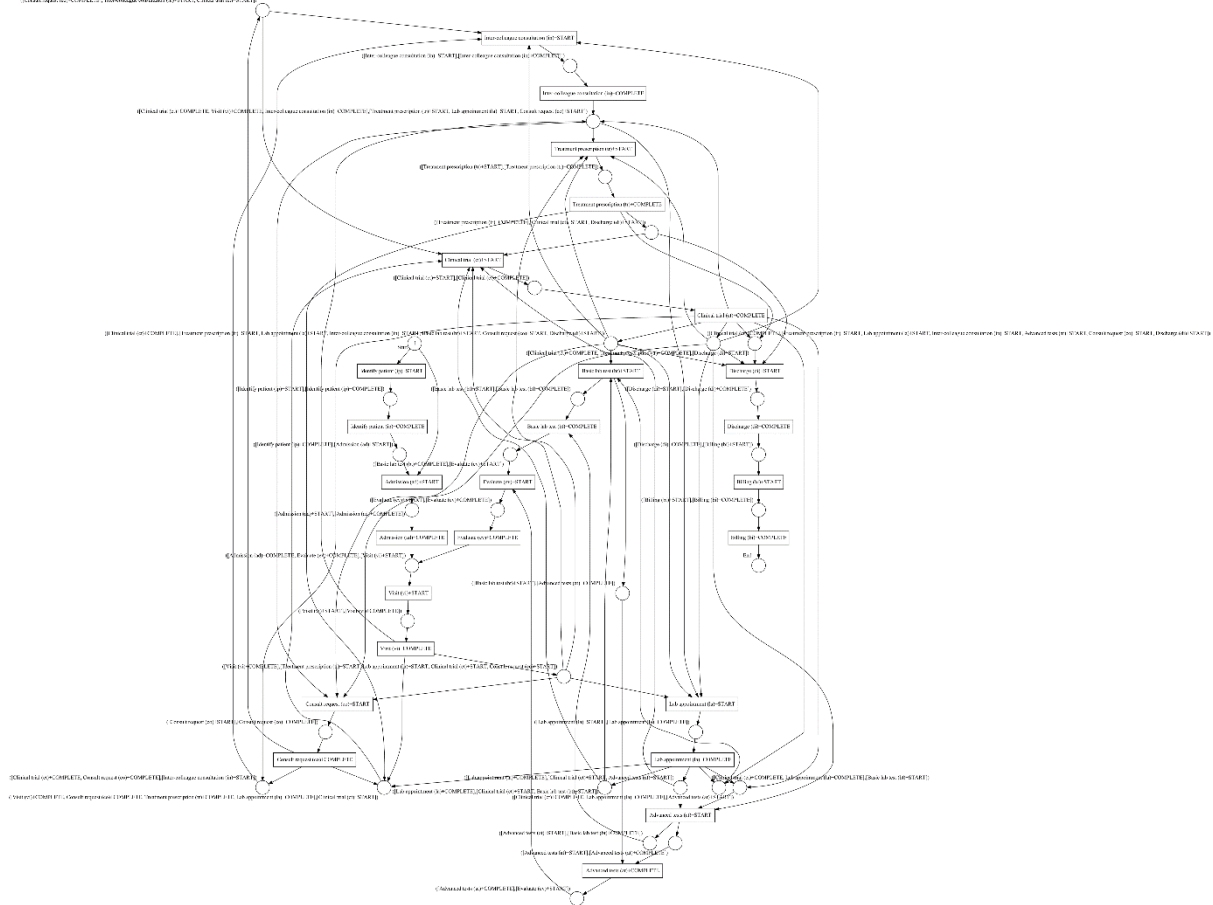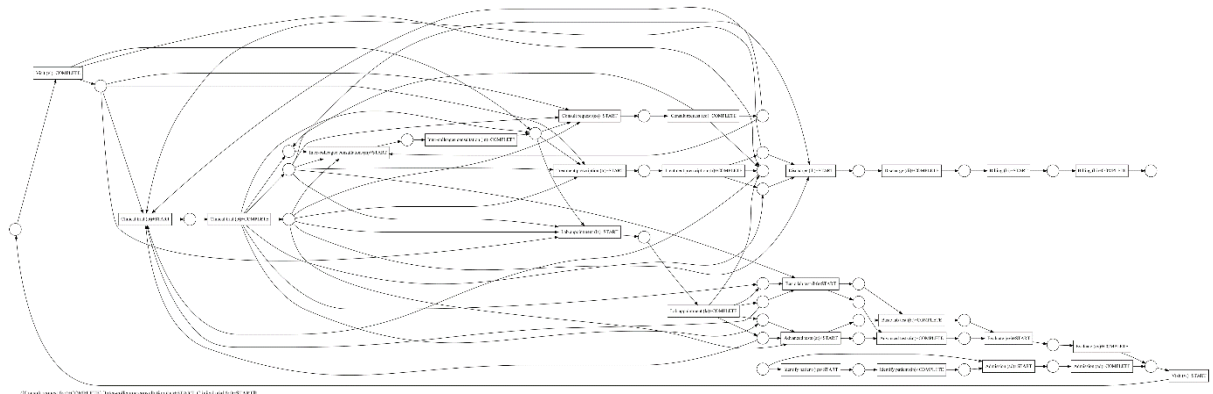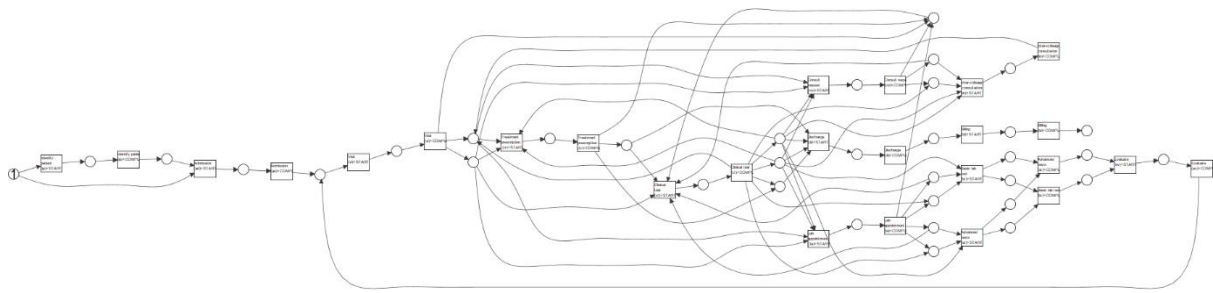
Alpha miner ++

Alpha# miner
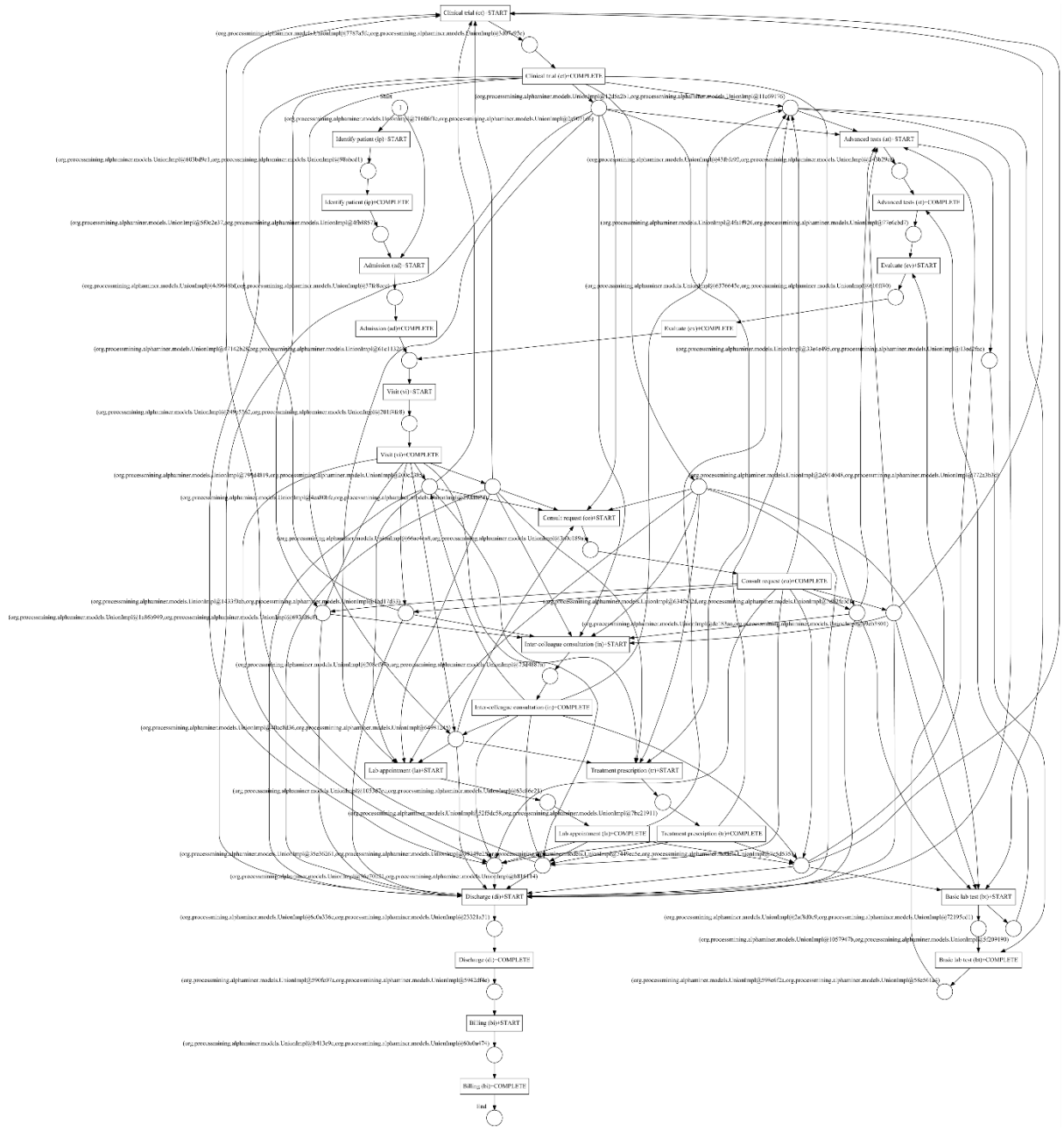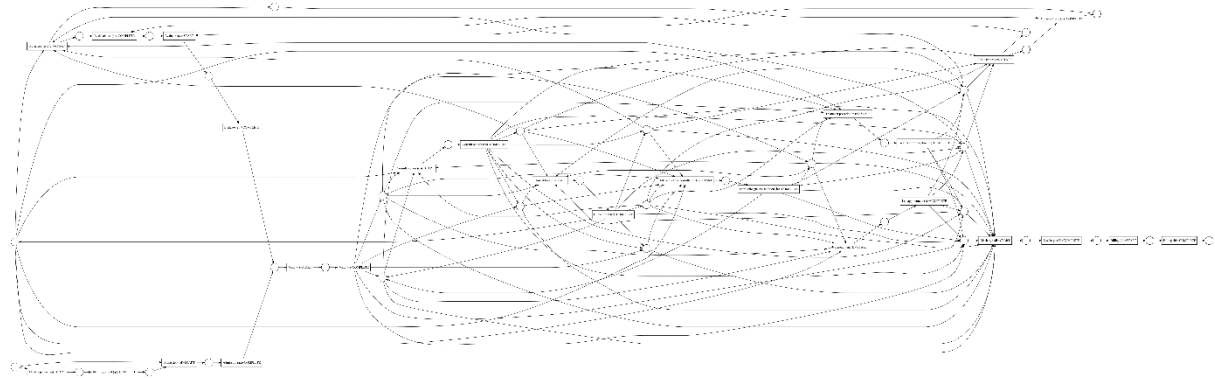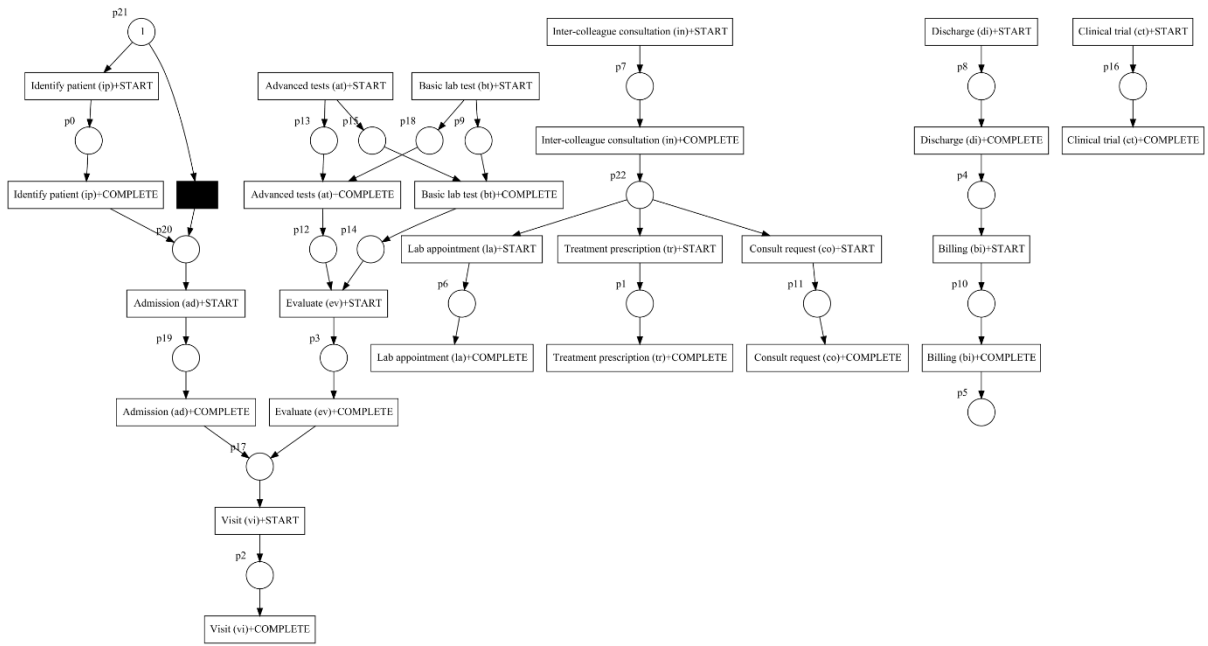
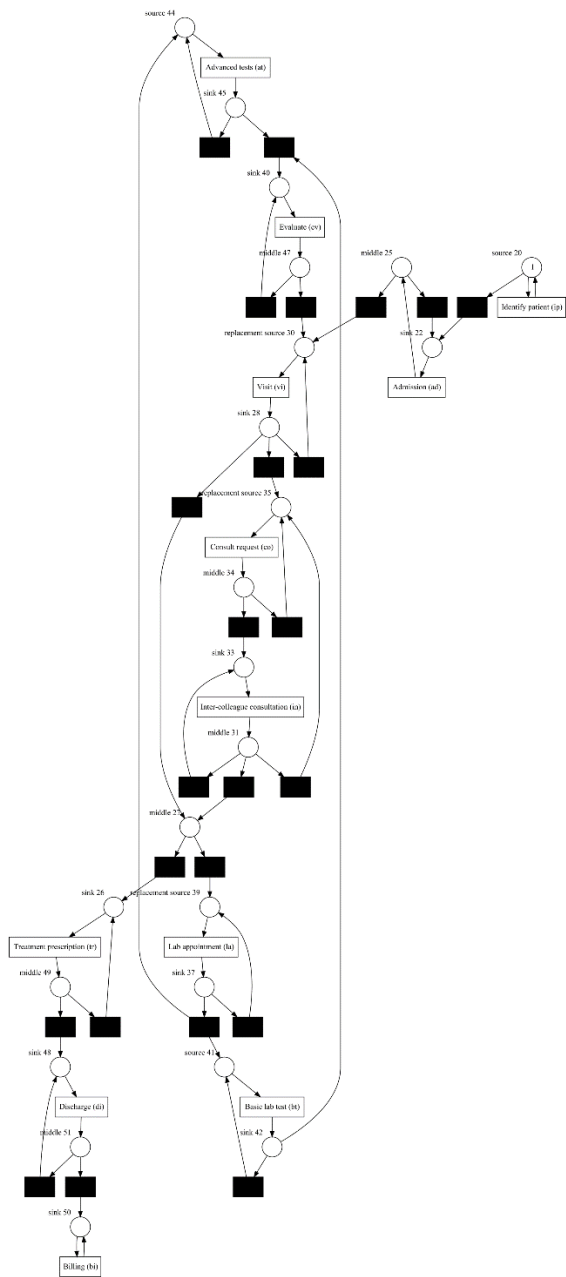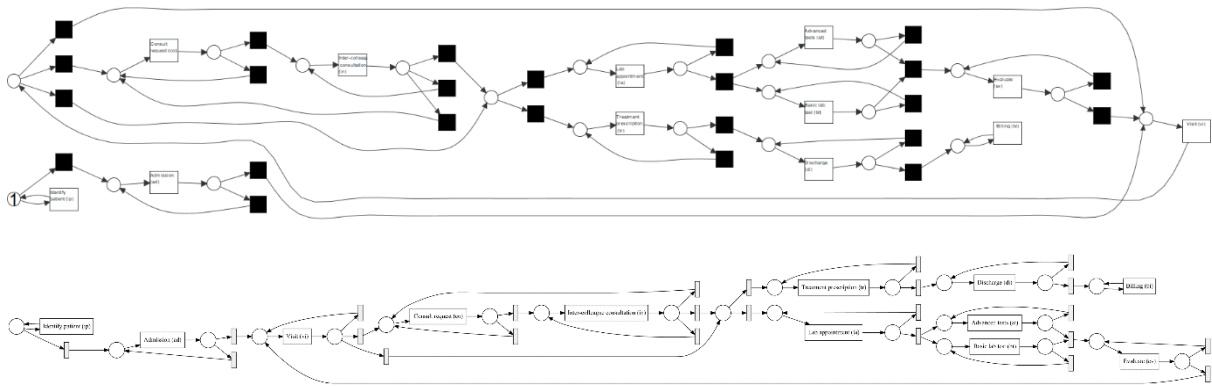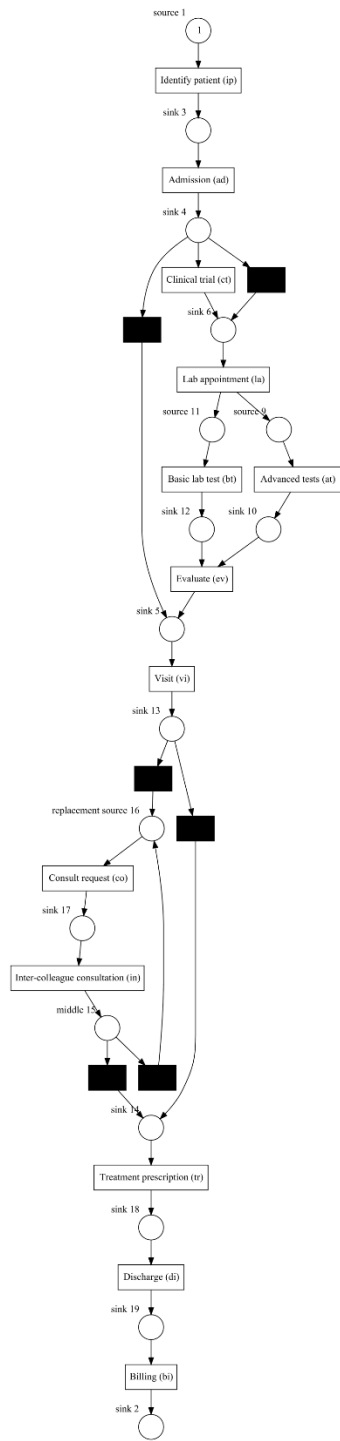Inductive miner with perfect log fitness

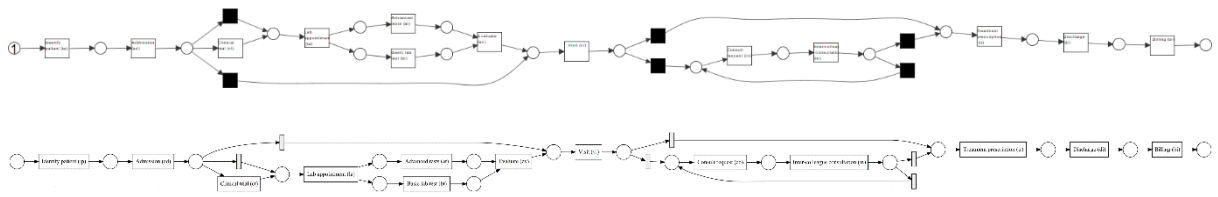Inductive miner with 0.01 Noise threshold

Inductive miner with 0.05 Noise threshold

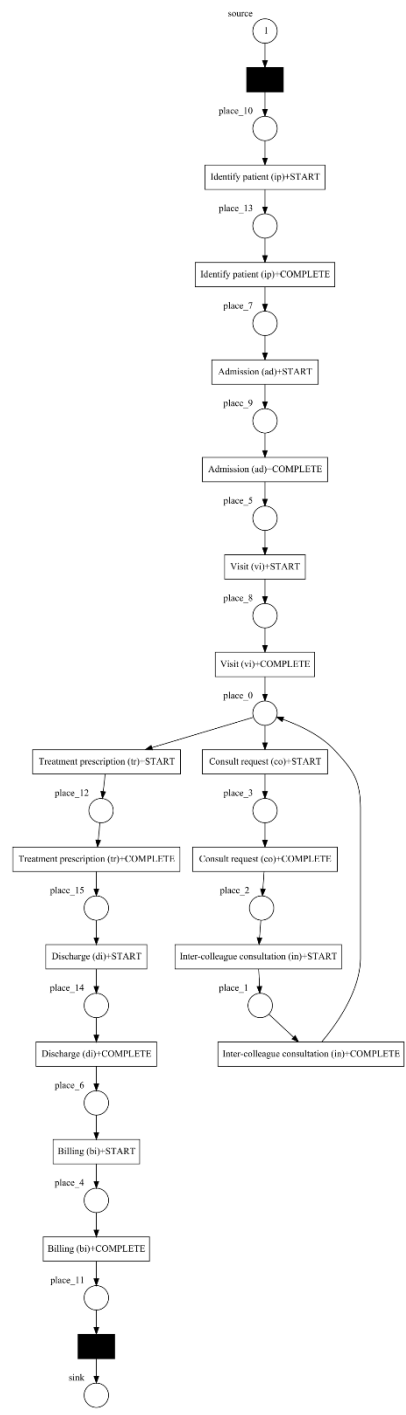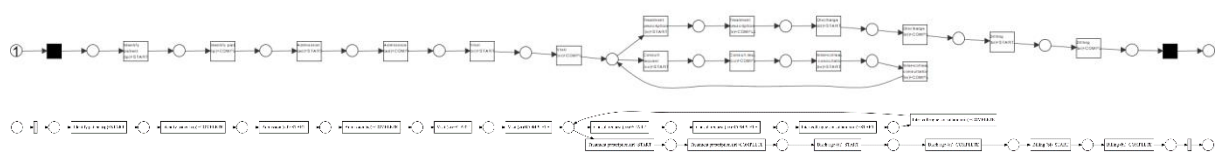# Inductive miner with 0.1 Noise threshold

Inductive miner with 0.2 Noise threshold

ILP-based process discovery (express)

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM)
COMPLETION REPORT - PART 1 OF 2
COURSEWORK REQUIREMENTS*

* NOTE: Scores on this Requirements Report reflect quiz completions at the time all requirements for the course were met. See list below for details. See separate Transcript Report for more recent quiz scores, including those on optional (supplemental) course elements.

- **Name:** Jelle Gerritsen (ID: 12850776)
- **Institution Affiliation:** Massachusetts Institute of Technology Affiliates (ID: 1912)
- **Institution Email:** j.w.s.gerritsen@student.utwente.nl
- **Institution Unit:** Student

- **Curriculum Group:** Human Research
- **Course Learner Group:** Data or Specimens Only Research
- **Stage:** Stage 1 - Basic Course

- **Record ID:** 59873965
- **Completion Date:** 19-Dec-2023
- **Expiration Date:** 19-Dec-2026
- **Minimum Passing:** 90
- **Reported Score*:** 92

REQUIRED AND ELECTIVE MODULES ONLY	DATE COMPLETED	SCORE
Belmont Report and Its Principles (ID: 1127)	19-Dec-2023	3/3 (100%)
History and Ethics of Human Subjects Research (ID: 498)	30-Nov-2023	4/5 (80%)
Basic Institutional Review Board (IRB) Regulations and Review Process (ID: 2)	17-Dec-2023	5/5 (100%)
Records-Based Research (ID: 5)	17-Dec-2023	4/4 (100%)
Genetic Research in Human Populations (ID: 6)	17-Dec-2023	4/5 (80%)
Populations in Research Requiring Additional Considerations and/or Protections (ID: 16680)	18-Dec-2023	5/5 (100%)
Research and HIPAA Privacy Protections (ID: 14)	19-Dec-2023	4/5 (80%)
Conflicts of Interest in Human Subjects Research (ID: 17464)	19-Dec-2023	5/5 (100%)
Massachusetts Institute of Technology (ID: 1290)	19-Dec-2023	No Quiz

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

Verify at: www.citiprogram.org/verify/?k65e390f0-a420-4c13-a016-84ad320b8a7c-59873965

Collaborative Institutional Training Initiative (CITI Program)
101 NE 3rd Avenue
Suite 320
Fort Lauderdale, FL 33301 US

Email: support@citiprogram.org
Phone: 888-529-5929
Web: https://www.citiprogram.org

COLLABORATIVE INSTITUTIONAL TRAINING INITIATIVE (CITI PROGRAM)
COMPLETION REPORT - PART 2 OF 2
COURSEWORK TRANSCRIPT**

** NOTE: Scores on this Transcript Report reflect the most current quiz completions, including quizzes on optional (supplemental) elements of the course. See list below for details. See separate Requirements Report for the reported scores at the time all requirements for the course were met.

- **Name:** Jelle Gerritsen (ID: 12850776)
- **Institution Affiliation:** Massachusetts Institute of Technology Affiliates (ID: 1912)
- **Institution Email:** j.w.s.gerritsen@student.utwente.nl
- **Institution Unit:** Student

- **Curriculum Group:** Human Research
- **Course Learner Group:** Data or Specimens Only Research
- **Stage:** Stage 1 - Basic Course

- **Record ID:** 59873965
- **Report Date:** 20-Dec-2023
- **Current Score**:** 92

REQUIRED, ELECTIVE, AND SUPPLEMENTAL MODULES	MOST RECENT	SCORE
History and Ethics of Human Subjects Research (ID: 498)	30-Nov-2023	4/5 (80%)
Basic Institutional Review Board (IRB) Regulations and Review Process (ID: 2)	17-Dec-2023	5/5 (100%)
Belmont Report and Its Principles (ID: 1127)	19-Dec-2023	3/3 (100%)
Records-Based Research (ID: 5)	17-Dec-2023	4/4 (100%)
Genetic Research in Human Populations (ID: 6)	17-Dec-2023	4/5 (80%)
Populations in Research Requiring Additional Considerations and/or Protections (ID: 16680)	18-Dec-2023	5/5 (100%)
Research and HIPAA Privacy Protections (ID: 14)	19-Dec-2023	4/5 (80%)
Conflicts of Interest in Human Subjects Research (ID: 17464)	19-Dec-2023	5/5 (100%)
Massachusetts Institute of Technology (ID: 1290)	19-Dec-2023	No Quiz

For this Report to be valid, the learner identified above must have had a valid affiliation with the CITI Program subscribing institution identified above or have been a paid Independent Learner.

Verify at: www.citiprogram.org/verify/?k65a390t0-a420-4c13-a016-84ad320b8a7c-59873965

Collaborative Institutional Training Initiative (CITI Program)
101 NE 3rd Avenue
Suite 320
Fort Lauderdale, FL 33301 US

Email: support@citiprogram.org
Phone: 888-529-5929
Web: https://www.citiprogram.org