Healthcare process analysis

validation and improvements of a data-based method using process mining and visual analytics

Torres Ramos, L.

Award date:
2009

Eindhoven, October 2009

Healthcare Process Analysis: validation and improvements of a data-based method using process mining and visual analytics

by Leonardo Torres Ramos

Student identity number 0641860

in partial fulfilment of the requirements for the degree of

Master of Science
in Operations Management and Logistics

Supervisors:
Dr. M.H. Jansen, TU/e
Dr. A.J.M.M. Weijters, TU/e

TUE. Department Technology Management.
Series Master Theses Operations Management and Logistics

Subject headings: process analysis, data analysis, health care, hospitals

EXECUTIVE SUMMARY

This graduation project is intended to help healthcare organizations in obtaining relevant process information for improvements in their clinical pathways. Due to the fact that currently healthcare organizations are trying to improve their care services with low costs, they are focusing in developing healthcare process analysis to obtain relevant information. In this context, this graduation project formulates the following problem statement:

Problem definition: Propose a method to analyze healthcare process-related data with the purpose of presenting useful process information for healthcare organizations. Additionally, this information should be easily obtained from the method and interactively offered so the users can utilize it for improvement projects.

In order to solve this problem, this graduation project proposes a validation of an existing method developed also for healthcare organizations (Riemers 2009). This method was designed with the same purpose of producing process-related information in healthcare environments. However, it has been applied only in one healthcare location and its validity should be assessed. Therefore, this graduation project applies the method in a different healthcare environment. The Academisch Medisch Centrum (AMC) in Amsterdam The Netherlands, offered a business case for this purpose. This business case contains process-related information about the gynecological oncology department of the AMC. This department represents a different treatment process compare to the one used by Riemers (2009) during the development of the method. Besides the validation process, this graduation project also attempts to improve and extent the existing method by adding useful analyses that produce understandable process-related information for healthcare stakeholders.

The used method in the validation process and analyses of this graduation project utilizes process mining and visual analytics techniques to get the results. The combination of these techniques supposes complementary results due to the fact that process mining offers process related information by looking at the inside of the processes while visual analytics can present this information in a clearer way. These characteristics are especially useful in a healthcare environment due to the inherent complexity in healthcare processes. The tools proposed in the method and used in this graduation project for the process mining and visual analytics analyses are the ProM and the MagnaView tools respectively. Therefore, these tools and techniques were also applied into the business case of the AMC.

The method developed by Riemers (2009) proposed seven different phases for such a project. These phases are:

1. Build database
2. Introduction session
3. Preliminary analysis
4. Preliminary meeting
5. 2nd analysis
6. Final meeting
7. Documentation

Each of these phases contains different steps and analyses to fulfill its goals. In the case of the validation process, all phases were followed as the method proposes. First, a database was constructed and loaded into the tools. Then, a meeting was planned in order to explain the benefits of this method to the different medical stakeholders. Additionally, these stakeholders provided the information requirements in order to start with the process analyses. After finishing the preliminary analysis phase, a second meeting was planned to show the obtained results and to get feedback on how to continue with the analysis and where to focus it. This feedback provided with the information required to perform the 2nd analysis phase. At the end of this phase, a final meeting was planned in order to present all the results. Finally, some documentation was elaborated so the medical actors can have the results available for improvement projects.

The conclusions of the validation process produced an evaluation on the steps and analyses of the method based on the feedback received during the validation process from the medical stakeholders. This evaluation judged the usefulness of each step and analysis in order to determine if they should remain in the method. Moreover, the results suggested that the method produced satisfactory results. However, there were some identified opportunities to improve it. Mainly, it was determined that these opportunities could be covered by extending the usage of the process mining tool due to the fact that the benefits of this tool has not been completely exploited yet. Therefore, some extra analysis was performed with this tool. This extra analysis was intended to repeat the analyses performed with the MagnaView tool (visual analytics tool) in the ProM tool (the process mining tool) in order to determine which results present more benefits for healthcare environments. Then, some criteria were defined in order to choose the best option for each analysis. The comparison based on the criteria produced a more extensive usage of the process mining tool.

The final part of this graduation project proposes a resulting method that takes into account the conclusions made from the validation process, the feedback received from the medical stakeholders of the AMC, and the results of the extra analysis. This new method proposes more analysis steps and it covers more aspects of a healthcare environment, like the organizational and process understanding aspects among the medical actors. This graduation project concludes by suggesting future research opportunities and a validation of the new analysis steps included in the method. Therefore, it can be said that this graduation project proposes a solution to the problem definition stated at the beginning. However, its applicability in different healthcare environments still has to be assessed.

CONTENTS

1. INTRODUCTION

This chapter describes the research area of this graduation project. First, the assignment description is elaborated in order to explain the general background of the project and to determine the main idea of this thesis. Second, the problem statement is described to specify the research questions and research goals. Third, the research approach is presented followed by the description of the business case applied in this graduation project. Finally, the outline of the remaining chapters is given.

1.1 Assignment description

Currently, healthcare organizations all over the world are increasing the usage of information systems and information technologies to improve their service. According to Anderson & Balas (2008), information technology has been recognized in many countries as an indispensable tool to improve patients care in healthcare organizations.

In The Netherlands, the increment in the usage of information technologies in hospital settings is being encouraged also by the necessity to reduce costs in the context of the new government regulations. Since 2005, Dutch government changed the method for paying the services delivered by hospitals. Nowadays, hospitals get their income through the so called DBC-code system (Diagnose Behandeling Combinaties). The DBC-codes are healthcare products that contain the entire steps needed for a certain treatment process. The DBC-codes are defined as a predefined average care product which a care provider selects based on the care demand of the patient (DBC web-site 2009). Furthermore, each DBC has a fixed price and the hospitals get paid only for the total DBC. In other words, hospitals do not get paid for each action (e.g. lab tests) but only for the complete DBC-code. Therefore, healthcare organizations are focusing on ways to perform a DBC as efficient as possible, with high quality care and low costs.

In order to manage DBC healthcare processes for improving the care service with low costs, information system technologies have been used as tools that help in accomplishing these goals. One of the benefits offered by information systems is the process information that can be obtained from the system if the goal is to achieve improvements. In healthcare organizations, treatment process improvement projects can be considered as Business Process Management (BPM) projects. According to Aalst et al. (2003), Business Process Management can be defined as: “Supporting business processes using methods, techniques, and software to design, enact, control, and analyze operational processes involving humans, organizations, applications, documents and other sources of information”. The BPM projects have a specific life-cycle. As is shown in Figure 1, the BPM project life-cycle identifies five phases: design, configuration, execution, control, and diagnosis, and it aims to support the whole process life-cycle (Aalst et al. 2007a).

Nowadays, Business Process Management systems provide support for the complete BPM life-cycle. Business Process Management systems are generic software systems, driven by explicit process designs to enact and manage operational business processes (Aalst et al. 2003). Currently, Business Process Management systems are also focusing on the area of Business Process Analysis (BPA). This area is mainly related with the diagnosis phase of the BPM life-cycle.

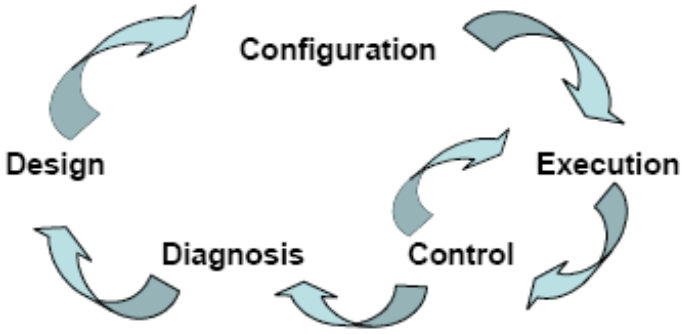


Figure 1: The BPM life-cycle

The area of Business Process Analysis includes approaches like process mining and visual analytics. In the efforts of healthcare organizations for obtaining relevant process knowledge for redesign and improvement projects, these approaches have proven to offer useful process information. Process mining and visual analytics are data-based approaches for process analysis that have been successfully applied in a hospital setting, in a previous graduation project at Eindhoven University of Technology. In that graduation project a data-based method for process analysis in healthcare using process mining and visual analytic tools was developed (Riemers 2009). The main idea of that method is that the combination of process mining and visual analytics could be a good option for getting relevant healthcare process information. These approaches appear to be good complements because process mining is a process driven approach which looks at the inside of the processes while visual analytics can help users to obtain clearer process insights. The tools used in the method for process mining and visual analytics are the ProM tool and the MagnaView tool respectively. Both approaches and tools are fully explained in Section 2.1 and Section 2.2 respectively.

The main objective of this graduation project is to validate the method developed by Riemers (2009) in a different healthcare environment due to the fact that it has been used only once and in one specific healthcare location. The goal of this validation is to establish if the method complies with the process information requirements of a healthcare organization for process improvements. For this purpose, this graduation project analyzes a case study of the Academisch Medisch Centrum (AMC) in Amsterdam, The Netherlands. Next section derives the problem statement and research goals of this graduation project.

1.2 Problem statement

Before any attempt to analyze healthcare processes using information technologies, it is important to highlight that healthcare processes are not simple processes, due to their inherent characteristics. According to Anyanwu et al. (2003), healthcare processes have three main characteristics:

- Firstly, they are complex, due to the fact that they involve clinical and administrative tasks, large volumes of data, patients and personnel.
- Secondly, they are dynamic, because changes in healthcare treatments, drugs and protocols are common.
- And finally, they are large-scale based, for the reason that their processes can involve several healthcare organizations and run over long periods of time.

A consequence of these characteristics is that "it is not known what happens in a healthcare process for a group of patients with the same diagnosis" (Mans et al. 2008a). Moreover, and as stated in the previous section, healthcare organizations are searching for reliable options that can present relevant process information about their healthcare processes. These options must include proved tools and analyses that can facilitate healthcare process analyses. In other words, the options for obtaining relevant healthcare process information must offer interactive and easily obtained results by means of predefined analyses that guarantee useful results. Therefore, the problem definition in this thesis is defined as:

Problem definition: Propose a method to analyze healthcare process-related data with the purpose of presenting useful process information for healthcare organizations. Additionally, this information should be easily obtained from the method and interactively offered so the users can utilize it for improvement projects.

As stated in the previous section, a method for healthcare process analysis has already been developed using process mining and visual analytics techniques. This method was developed by Riemers (2009) in a previous graduation project. Furthermore, the AMC agreed to provide a business case in order to investigate if the method complies with offering relevant process information. Therefore, the first and main goal of this graduation project is defined as follows:

Research goal 1: Validate the method developed by Riemers (2009) in the healthcare environment of the AMC.

Healthcare organizations have different stakeholders with different backgrounds that need information in different specific ways. To overcome this difficulty, it must be determined which stakeholders could be the users of the process analysis results. Because healthcare organizations are looking for process information that can contribute to the improvement of patient treatments and the standardization of its processes, the stakeholders chosen as the users of the results of this method should be capable of producing healthcare process improvements. To do this, the chosen stakeholders must completely understand the results and their applicability in healthcare process improvements. This leads to our second research goal:

Research goal 2: Determine the users of the analysis results and if these users find the results understandable.

After the validation process, it could be determined if the process analysis results of the method complied with the information requirements of the AMC, in order to establish if some adjustments or extensions need to be made to the method (the initial information requirements of the AMC are fully explained in Section 3.2). Therefore, this analysis has to take into account the possibility to improve the method by extending the combination or the individual utilization of the used approaches (process mining and visual analytics). Therefore, the third goal of this graduation project is defined as follows:

Research goal 3: Analyze if the developed method needs to be partially changed or improved in order to produce more relevant healthcare process information.

This graduation project tries to achieve these research goals in the remaining chapters. The project will investigate these goals by means of a business case, the validation process, and research in the process mining and visual analytics tools. Next section describes the research approach used for each of the research goals established in this section.

1.3 Research approach

The research goals previously established serve as a guideline to perform this graduation project. However, in order to achieve these goals some research approach has to be defined for each research goal.

The first and main goal mentioned in the last section establishes that a validation process of an existing method for healthcare process analysis has to be carried out. According to Balls et al. (1995), a validation is the process by which the reliability and relevance of a procedure are established for a specific purpose. In this case, reliability describes whether a procedure can be performed reproducibly among time while relevance describes whether a procedure is meaningful and useful for a particular purpose. The authors also highlight that several approaches to validation may be scientifically acceptable, depending on the particular purpose and goal of the study. Therefore, the term validation in this graduation project is defined as the process of checking the reliability and relevance of the method in satisfying the process information needs for healthcare process improvements of the different stakeholders at the AMC.

As is explained in Section 2.3, the existing method proposed different phases for a healthcare process analysis. These phases are going to be followed in this validation process as proposed in the method in order to check its reliability. Moreover, in the analysis phases, the method proposes certain views and analyses to get the results. The views and analyses that are going to be used during this validation project are the ones that comply with the following characteristic:

- They must contribute to fulfill the requirements of process information for the different stakeholders at the AMC. In other words, they must be relevant.

The compliance of this characteristic, by the used views and analyses, is going to be judged with the process-related questions and feedback received from the stakeholders during the analysis phases of the validation process. Thus, their requirements and feedback are going to determine if the information offered by the different views and analyses is useful. By following this approach in the validation process, the method will be validated according to its main goal which is to cover the necessities of process related information in a healthcare organization. Furthermore, the analyses will not be restricted to the analyses proposed in the method, and in that way the method could be extended or improved if during the project it is shown that some analyses offer better results. Chapter 3 and the first part in Chapter 4 explain the validation process and the validation results under this perspective.

The second research goal of this graduation project states that the process analysis results should be understood by the different healthcare stakeholders. The fulfillment of this research goal is going to be judged with the feedback received during the final meeting at the AMC where all the obtained results are presented to the different stakeholders in the hospital. This meeting is part of the phases proposed in the method, and it serves to establish in what extent the initial goals of the analysis were met. The second part

in Chapter 4 describes the feedback received during this meeting. That section uses the feedback to establish if this second research goal was reached.

Finally, the third research goal establishes that an extra analysis has to be carried out in order to determine, from the feedback of the different stakeholders at the AMC, if some additional investigation is needed with the process mining and visual analytics tools to extent or to improve the existing method. Certainly, this extra analysis is based on the information requirements of the AMC. If the requirements are fulfilled with the method then no extra analysis is needed but if some opportunities can be identified, then the investigation on the tools must be completed in order to realize these opportunities. If needed, the approach that will be used to accomplish this goal is to develop all the analysis using both tools (the ProM tool and the MagnaView tool) in order to compare the advantages and disadvantages offered by both analyses, and to choose which one represents the best option for a healthcare environment. First, the analyses proposed in the method are repeated with both tools. Then, the extra analyses and tools developed during the project would also be repeated in both tools in order to establish which one represents the best option. Chapter 5 presents the most important findings during these analyses. Additionally, Chapter 5 also describes the criteria used to choose the results between the options offered by the process mining and the visual analytics tools.

1.4 Business case

In order to develop a solution for the problem statement defined in Section 1.2, this graduation project analyzes a case study in the Academisch Medisch Centrum (AMC) in Amsterdam, The Netherlands. The AMC hospital is located in Amsterdam, in the south-eastern part of the city. The AMC officially has 1,002 beds, each year 25,000 patients are admitted, and there are 35,000 day admissions and over 350,000 outpatient visits. The AMC-complex houses the university hospital and the medical faculty of the University of Amsterdam, as well the Netherlands Institute for Neuroscience and the medical department of the Royal Tropical Institute. Also a number of biotech companies – partly AMC spin-offs – are located on the premise (AMC web-site 2009).

In the recent years, the AMC has been performing projects for customer-focused and cost-effective care in order to improve their service, for example, reducing waiting times. Some of these projects have been well documented. According to Elkhuizen et al. (2007), two redesign projects were performed at the AMC for a specific patient group. The goals of these projects were to improve patient care processes for the gynecological oncology and dyspnea patient groups. The approach used to get information for these redesign projects included activities like interviewing people, observations and checking patient documents. Typically, these activities are very time consuming. Therefore, the AMC identified that a different type of project can be performed. This new project showed the applicability of process mining in the healthcare environment (Mans et al. 2008a). This project offered the opportunity to obtain insights in the healthcare processes and perform an analysis in a quicker way, even before going to the department or meeting the people involved in the process, saving a lot of time in collecting the data. Clearly, process mining showed to be a useful tool in the healthcare analysis process. However, the process results obtained from this analysis showed the complexity and variation of the analyzed processes. Furthermore, the medical specialists found it difficult to understand, interpret and use the process-related results of the analysis to improve the healthcare processes. Therefore, and as visual analytics has already shown its

benefits in presenting clearer process information for healthcare specialists, the AMC became the perfect scenario to comply with the research goals of this graduation project.

The groups of patients included in this business case are from the gynecological oncology department. Thus, the analyzed patients are patients that require a treatment of cancer. The data used to answer the problem statement comes from a file that contains 682 patients and 43,615 events performed at the AMC for this group of patients. These patients were in the diagnosis and treatment phases of the process between the 3rd of January 2005 and the 20th of March 2008. Specifically, the data includes six different DBC-codes: M11 "maligniteit vulva", M12 "maligniteit vagina", M13 "maligniteit cervix", M14 "maligniteit endometrium", M15 "maligniteit myometrium" and M16 "maligniteit ovarium / tuba".

The analyzed file that contains the raw data used in this project comes from the billing information system of the AMC. As explained in Mans et al. (2008a), despite hospitals may have different IT applications, the information contained in the billing systems of hospitals is process-related because they have to guarantee that the hospital gets paid for all the delivered activities to the patients. The disadvantage is that the timestamps of the activities were only "days". In other words, we do not know the exact order of activities per day per patient.

Next section describes the outline planned for the complete project and it explains how this report is organized.

1.5 Outline

This chapter has presented the background information and the motivation for this research project. The research goals and the research approach were also described. The remainder of this thesis is structured as follows: Chapter 2 presents the preliminaries about the different approaches that are used during this graduation project. This chapter explains all the details about process mining and visual analytics. Additionally, it describes the characteristics of the tools used for process mining and visual analytics, the ProM tool and the MagnaView tool respectively. This knowledge is indispensable for a full comprehension of the remaining chapters. Additionally, it also explains the goals, phases, content, and main conclusions of the developed method for healthcare process analysis that uses those process mining and visual analytics tools. Chapter 3 describes the validation process at the AMC. It describes all the steps performed when applying the method at the AMC and all the results obtained from the analysis. The focus of this chapter is on the first research goal of this project and it describes all the phases of the validation process. In Chapter 4, the focus is on the conclusions of the validation process and this analysis is divided in two main parts. First, the analyses phases proposed in the method are evaluated from a technical point of view. And second, the opinion of the stakeholders about the method is presented focusing mainly on the understandability of the results as explained in the second research goal of this project. Chapter 5 describes possible improvements of the method using the process mining and visual analytics tools. These improvements take into account the conclusions drawn in the previous chapters. The main idea of Chapter 5 is to cover the third and final research goal, explained in Section 1.2. In Chapter 6, the resulting method is explained, followed by Chapter 7 where the main conclusions, limitations of this project and future research recommendations are described in detail.

2. PRELIMINARIES

As is mentioned in the previous chapter, process mining and visual analytics have been successfully applied in a hospital setting to obtain relevant process information. These approaches seem to be good complements when applied in healthcare environments due to the fact that the main purpose of process mining is to look inside of the processes while visual analytics focuses on presenting this process information in a clearer way. A previous developed method (Riemers 2009) proposes a way to combine these tools for a process analysis project in a healthcare environment. This chapter presents the preliminary knowledge of process mining, visual analytics and the existing method of healthcare process analysis. It gives a short explanation of what can be expected from each approach, a general definition of the tools used for process mining and visual analytics, and an explanation of the development, content and conclusions of the method for healthcare analysis as proposed in Riemers (2009).

2.1 Process mining

Process mining deals with the problem of limited information about what is really happening in the processes (Mans et al. 2008a). This information is of course needed if the goal is to achieve process improvements. The aim of process mining is "the automatic construction of models explaining the behavior observed in the event logs" (Aalst et al. 2007b). Process mining offers techniques to discover processes and data, and to determine organizational and social structures from the event logs. The event logs are the files that contain information about process instances and its contexts (Aalst et al. 2007b). Thus, process mining works with real data, the one that comes directly from the information system of the organization, in order to discover, monitor and improve real processes (Mans et al. 2008a).

There have been some documented examples of the successful application of process mining in the service industry (Aalst et al. 2007b). Additionally, process mining has been already used in healthcare for different purposes. For example, to discover patterns of process execution (Lin et al. 2001), to get insights for improvements in care flows (Mans et al. 2008a) and even to discover fraudulent and abusive behaviors in healthcare organizations (Yang & Hwang 2006). Process mining can be conducted by a process mining tool called ProM. The ProM tool is an independent platform that supports a wide variety of process mining and data mining techniques (Mans et al. 2008b).

Next sections provide an overview of process mining and more detailed information about the tool used to conduct the process mining analyses in this graduation project, the ProM framework.

2.1.1 Overview of process mining

Process mining strives to deliver process information about what actually happens in an organization. This organizational reality is extracted from the event logs. The event logs are the files that have a sequential record of events such that each event refers to an activity. Moreover, these event logs can have additional information such as the performer and originator of the event, the timestamp of the event, or data elements recorded with the event (Aalst et al. 2008). Therefore, the formal requirements of an event log in order to serve as input to process mining are (Aalst et al. 2007b):

i. Each event refers to an activity (i.e. a well defined step in the process)

ii. Each event refers to a case (i.e. a process instance)
iii. Each event can have a performer also referred to as the originator (i.e. the person executing or initiating the activity)
iv. Each event refers to a timestamp (i.e. the time at which it was recorded)
v. Events are totally ordered.

With this information as input, process mining attempts to construct process models explaining the behavior found in the event logs. As was mentioned before, process mining addresses the problem of limited information about what is actually happening in an organization.

As is shown in Figure 2, there are three basic types of process mining:

- Discovery: This type of process mining appears when there is no a-priori model and based on the information of an event log a process model is constructed.
- Conformance: This type of process mining uses an a-priori model to confirm if this model conforms to the reality shown in the model extracted from the event log.
- Extension: This type also uses an a-priori model. This model is extended with a new aspect or perspective in order to enrich the model with the data contained in the event log.

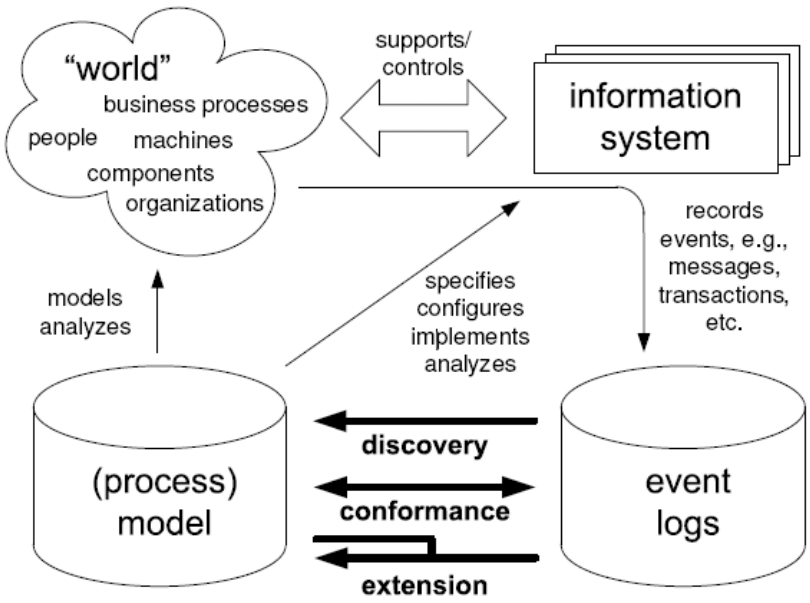


Figure 2: Three types of process mining: discovery, conformance and extension.

Nowadays, process mining tools are becoming available and they are being integrated into larger information systems (Aalst et al. 2008). The ProM framework is a powerful collection of process mining tools that provides an extensive set of analysis techniques which can be applied to real processes and it supports all three types of process mining (Aalst et al. 2007c). The ProM framework is the tool used to get the process mining results of this graduation project. Therefore, next section explains this tool with more detail.

2.1.2 The ProM framework

The ProM framework was developed by a group of researchers at Eindhoven University of Technology. The main idea of this framework is to bring together different process mining tools, all into one integrated environment. ProM is open source and uses a plug-able architecture where people can add new process mining techniques by adding plug-ins into the framework (Aalst et al. 2007c). A plug-in is a piece of software or functionality that can be added or removed without affecting the functionality of the framework.

In 2004 the first version of ProM was made public. Since then, ProM has been extended dramatically and currently dozens of researchers are developing plug-ins for ProM (Aalst et al 2007c). The ProM framework uses log files in the standard Mining XML (MXML) log format. In order to convert a log into a MXML format, a ProM import framework has also been developed (Gunther & Aalst 2006). This tool converts logs from various types of information systems into a MXML format.

The ProM framework includes five different types of plug-ins:

- Mining plug-ins which implement algorithms that mine models from the event logs.
- Analysis plug-ins which typically implement some property analysis on some mining result.
- Conversion plug-ins which implement conversions between different data formats.
- Export plug-ins which make possible to export objects to a certain data format.
- Import plug-ins which make possible to import objects of a certain data format.

The ProM software version used in this graduation project was version 5.0. Next section continues with the description of the second technique used in this graduation project: visual analytics. First a definition and a general explanation are given followed by the description of the tool utilized in the process analyses of this project.

2.2 Visual analytics

Despite the benefits offered by process mining, this tool may not be enough in a healthcare environment. This is due to the fact that although ProM models reflect the reality of the processes, these models are typically too complex in healthcare analysis because of the variety of activities in healthcare processes. Moreover, healthcare process mining models are usually called spaghetti-like models due to the variations and great number of activities performed. Additionally, it is necessary to be a process mining specialist in order to interpret these models. Healthcare specialists, practitioners and decision makers in healthcare organizations usually do not have this background. Therefore, for these stakeholders it could be difficult to really comprehend process mining results.

As was mentioned before, visual analytics can help users to better understand process performance. In our everyday life we commonly use visual analytics tools. An illustration can be seen in a weather forecast, which depends on large amounts of data collected by different sensors, and that can be shown as a visual representation of the landscape that facilitates the process to comprehend the dynamics and patterns of the weather (Simoff et al. 2008). Visualization of large data sets can be used in the same way in a professional environment. For example, in the healthcare environment it could be possible to visualize

where the bottlenecks are in an organization, to see process patterns about certain processes, or to analyze the compliance of certain service levels specified by a hospital. Here, visualization will require high volume data collection, processing, mining, modeling, and communicating the models quickly to the decisions makers (Simoff et al. 2008). The emerging field of visual analytics "focuses on handling massive, heterogeneous, and dynamic volumes of information through integration of human judgment by means of visual representation and interaction techniques in the analysis process" (Keim et al. 2006). Consequently, visual analytics can represent a solution for presenting what is really happening in the healthcare processes in an understandable way to the healthcare specialists. A possible tool for visualization is MagnaView. This is the tool proposed in Riemers (2009) for the visual analytics part of the existing method for healthcare process analysis. This software is used to visualize data due to the fact that it "delivers innovative solutions to interactively analyze and visualize" data (MagnaView web-site 2009).

Next sections present first a general overview of visual analytics that highlight its main characteristics and finally, the description of the tool used for the analyses in this graduation project: MagnaView.

2.2.1 Overview of visual analytics

According to Thomas & Cook (2006), visual analytics is "the science of analytical reasoning facilitated by interactive visual interfaces". The authors highlight that the use of visual analytics tools and techniques must serve to:

- Synthesize information and derive insights from massive, dynamic, ambiguous and often conflicting data
- Detect the expected and discover the unexpected
- Provide timely, defensible, and understandable assessments
- Communicate assessment effectively for action.

Furthermore, the authors emphasize the importance in the quality of the data in order to produce good visualizations which are an essential aid to the analytical reasoning process. Good quality visualizations must comply with the following characteristics:

- Facilitate understanding of data
- Provide frameworks for analyzing special and temporal data
- Support the understanding of incomplete or misleading information
- Provide representations that enable full situation awareness while at the same time supporting development of detailed actions
- Support multiple levels of data and information abstraction, including integration of different types of information into a single representation.

Thus, visual analytics tools must represent the data that is suitable for analyses, capturing the important content in the information. The MagnaView tool was identified as a tool that complies with these characteristics. Additionally, it has been successfully applied in healthcare environments. Next section shortly describes the tool and its main characteristics.

2.2.2 MagnaView

The MagnaView tool focuses on the construction of visualizations. This tool can be used to visualize processes, specific cases or aggregated groups. As is shown in Figure 3, the main characteristic of this tool is its data analysis functionality. MagnaView does not use mining algorithms to visualize data but it leaves mining to the user. In other words, the tool can present the information of large datasets in only one visual representation, and from there, the users can work with this information in order to interactively select and analyze the data.

Using the tool, the data can be selected directly from the visualization, the users can create filters for unnecessary data, and users can zoom in the visualizations according to the desired detail in certain characteristic of the data. The MagnaView software version used in this graduation project was version 4.0.

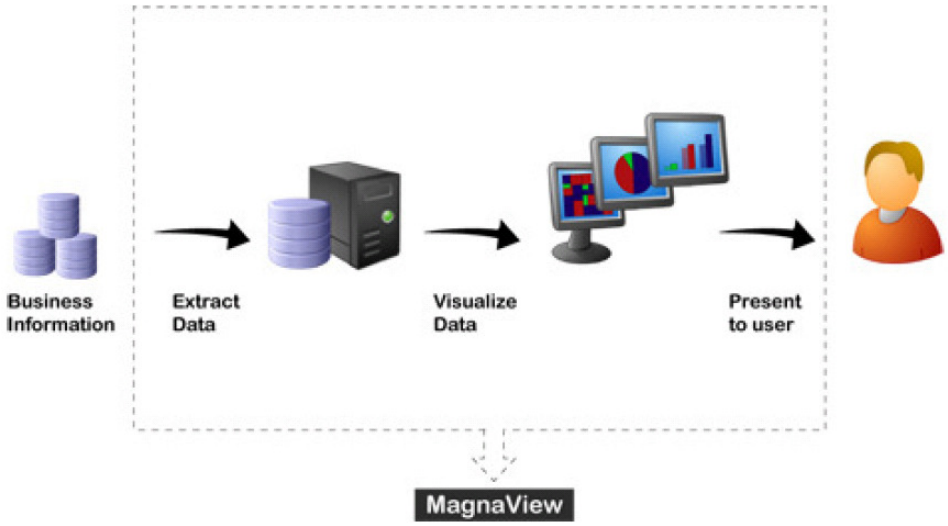


Figure 3: Data analysis process in MagnaView.

This chapter has presented the basic information regarding process mining and visual analytics as well as the description of the tools used in this graduation project for both techniques. Next section continues with the explanation and description of the developed method for healthcare analysis that uses the combination of ProM and MagnaView to get healthcare process information.

2.3 The existing method for healthcare process analysis

As was mentioned before, this method uses the combination of process mining and visual analytics tools in order to obtain process knowledge from healthcare environments. The method was developed in a business case analysis at a large hospital in the south of The Netherlands (Riemers 2009). The method proposes the use of process mining and visual analytics tools as the techniques for getting useful healthcare process information. Its main goal is to offer an alternative for process analysis in healthcare organizations. Next section continues with the description of the factors applied to justify the usage of both tools.

2.3.1 Success factors

According to Riemers 2009, some success factors were established as the factors needed to comply with the goals of the method. First, at the beginning of the project, when the analysts have not worked with the tools yet, they established the following factors:

- Results should be presented within limited time
- Process models should have a high fitness value
- The approach should be positively evaluated by the medical specialists and managers

Then, after working in the healthcare environment with the medical specialists and the process mining and visual analytics tools, the author added the following factors as a complement of the previous ones. These factors were derived by noticing that these characteristics are indispensable in order to achieve the goals of the method:

- The results should be simple to understand for the medical specialists and managers
- Interactive analysis should be possible
- The analysis should focus on certain aspects of the treatment process

The author stated that process mining and visual analytics do not comply with all these success factors, if the tools were used individually. Moreover, he affirmed that an analysis by means of these approaches was insufficient to actually have an impact. Therefore, he stated that, being the combination of both tools achievable, a method needed to be designed. In the following section this method is described.

2.3.2 Phases and content of the method

The sequential phases of the method are:

1. Build database
2. Introduction session
3. Preliminary analysis
4. Preliminary meeting
5. 2nd analysis (with adjustments)
6. Final meeting
7. Documentation

It is important to mention that in the analysis phases (phases 3 and 5 of the previous list), the author suggests the usage of certain plug-ins for the process mining analysis in ProM. Additionally, the author developed some views that were used in the visual analysis using MagnaView. These plug-ins and views were suggested as specific steps in these analysis phases. APPENDIX A shows the method in detail, with all the steps for each phase, as was proposed in Riemers (2009).

As was mentioned before, the method uses results from process mining and visual analytics tools. Therefore, the phases include actions using both approaches. The goal of the first phase, building the database, is to obtain a database which can be used to visualize and analyze treatment processes. This phase includes actions like collecting useful data, resolving the problems that may arise with missing data or inconsistent data, transforming the data into the needed formats for the analysis, and loading the data into the tools. The second phase proposes an introduction session with the healthcare actors to obtain a clear view on what is expected from the analysis. The actions here are tended to meet the specialists,

determine users, goals and important Key Performance Indicators (KPIs). Then, the preliminary analysis is carried out. The method recommends some analysis using the ProM and MagnaView tools to get useful results. Once the analysis is finished, the fourth phase states a preliminary meeting which should be held only with a small group of actors. The intention of this meeting is to inform the first results and to receive feedback on the current analysis. This session must determine follow-up activities, so the 2nd analysis phase can be developed. This phase includes actions similar to the ones proposed in the preliminary analysis phase but it should be adjusted according to the received feedback. The sixth phase is a final meeting with the entire team. In this meeting the results must be presented and the reached goals must be determined. Finally, some documentation is suggested so the actors can read the results back and perform a different analysis. Two files must be created by the analyst. A project document, in a text-file, which contains a clear description of the data and analyses performed, and a project file that could be used interactively.

The final part of this section elaborates on the main findings and conclusions in the development of the method.

2.3.3 Findings and conclusions in the development of the method

The main conclusions after the development of the method were that:

- The suggested combination of the techniques (process mining and visual analytics) had a positive effect, due to the fact that medical specialists were able to understand the results in an easier way.
- The method fulfilled all the success factors established in Section 2.3.1.
- The emphasis of the method was on the diagnosis phase of the BPM life-cycle. The designed method focuses on tools which could provide information about the structure of treatment processes.

The method proved, as a proof of concept, to be a good option for healthcare organizations in gaining information about their processes. Moreover, it was positively evaluated by the actors involved during the business case. With this tool developed and ready to be used, next section will elaborate on the validation process of the method performed at the AMC.

3. THE VALIDATION PROCESS

The goal of this research project includes recommending an approach for process analysis in the healthcare environment. To do so, a validation of a developed method for process analysis (Riemers 2009) was carried out. The method, explained in detail in Section 2.3, was developed at a large hospital in the south of The Netherlands. During the development of this method, two treatment processes of the hospital were analyzed: mamma care (cancer treatment) and diabetes foot. Therefore, this validation process is intended to examine if the method can offer useful process information to healthcare stakeholders in a different healthcare environment using the data from the gynecological oncology department of the AMC.

As was mentioned in Section 1.3, the research approach for this validation process is to check the reliability and relevance of the method in satisfying the process information needs of the different stakeholders at the AMC by using the phases, tools and views that contribute to fulfill those process information needs. The tools and steps for getting process information proposed in the method were taken into account in this project. However, some of these tools and steps were not followed in this project and the reasons to do so for each specific case are explained in the following sections. Additionally, and as proposed in the method, the healthcare stakeholders chosen to be the users of the results are medical specialists of the department and process managers due to the fact that these stakeholders are mainly responsible of the clinical pathways and they can implement and control improvement projects based on process analysis results. Thus, this method is intended to be validated by both stakeholders.

The next sections are dedicated to the seven phases of the developed method (build database, introduction session, preliminary analysis, preliminary meeting, 2nd analysis, final meeting and documentation). Each section presents an overview of the most important events and findings of this validation process at the AMC. Moreover, each section presents all the steps and actions done during this validation process. The order of the phases is sequential. The method starts with building the database for the tools and ends with the development of the final documentation that contains all the analysis results of this project. Hence, next section starts with the first phase of the method.

3.1 Build database

Mans et al. 2008a documented a previous process mining project performed also at the AMC. Some of the data used in that previous project were also included for the analysis of this graduation project. Therefore, the contact with the data manager at the AMC was already done.

The goal of this project was to analyze process-related information from the gynecological oncology department of the AMC. For that reason, the extracted data contained information on: patient IDs, name of the activity, involved department, timestamp for each activity and the DBC-codes (Diagnose Behandel Combinatie). Additionally, certain attributes were also included in the dataset for further analyses, like the age of patients. As was explained in previous sections, the data were extracted from the billing system of the hospital. The billing information system represented the best option to obtain process-related information for the reason that all the activities performed for each patient must be trustworthy acquired in order to charge the correct amount of money to each patient. However, it is obvious to notice that this

database includes administrative activities because its main purpose is to charge the health services to the patients. Therefore, some pre-processing activities led to identify and filter out these administrative activities from the dataset so only the process-related information kept included in the analyses. These pre-processing comprehended the actions of checking all the activities included in the dataset in order to identify the administrative activities. APPENDIX B shows in Table 10 the recognized administrative activities and in Table 11 the complete list of activities included in the analyses after filtering the administrative activities.

The next step was the transformation of the data into the right formats for the MagnaView tool and the ProM tool. MagnaView uses the MVN-format which was developed at MagnaView B.V. The data was extracted in a text file from the AMC billing information system and then exported into the tool to convert the file into a MVN-format. Later on in the project, the raw data in the text file was exported to create a XLS file from it because this file can be read by Microsoft Excel, a spreadsheet program that facilitates some pre-processing activities like adding attributes to the dataset for the analysis. Moreover, MagnaView tool can import data from TXT, XLS, MDB or ODBC among other formats. The log-file needed for the ProM tool was obtained directly from the AMC because this file was already produced in earlier performed analysis. Thus, the use of the ProM import tool was not required.

Finally, the data was loaded in the MagnaView and ProM tools for the analyses. According to Riemers (2009), the method establishes some activities for the building database phase in the method. These activities included the extraction, transformation and loading of the data into the tools. During this validation process, some of these activities were not done due to different factors. Next paragraphs mention which activities were not followed explaining the reasons to do so in each specific case:

- As explained at the beginning of this section, the contact with the data manager was already done because the data had already been identified before the start of this graduation project. However, this activity should be included in the method because it is necessary to ask for the information to the right person in a healthcare organization.
- The integration and consolidation of the data include actions in solving problems related with joining data from different systems. During the development of this method this activity was necessary because the data came from different information systems. However, because in this graduation project the data was extracted completely from the billing system of the hospital, this step did not require any action. Thus, as the information from the billing system has proved to be a good option of getting process-related information from the hospitals, this information system could be formalized as the source of information for this type of analysis. If the data can be extracted from the billing system of the hospitals then the integration and consolidation of data activity could be removed from the method.
- Finally, the usage of the ProM import tool was not necessary in our case because we already have the log-file in the right format. Nevertheless, this activity is essential to convert the data into the MXML-format.

To see a list of which activities of the method were done during this validation process at the AMC for the building database phase, go to APPENDIX C Table 12.

Once the data was ready for the analysis in both tools, the introduction session could be held. Next section elaborates on the details of that presentation.

3.2 Introduction session

The aim of the introduction session is to present the benefits of the analysis to the actors in the hospital. Additionally, they can obtain a clear idea of what to expect from the project. This session was performed at the AMC and the group of analysts met with some of the actors in the hospital. The method proposes to include different stakeholders in the project, like managers and medical specialists. Therefore, the presentation was shown first to the lead manager of the process-innovation department and then to the lead specialist from the gynecological-oncology department. As established before, these actors were determined as the users of the analysis results.

The status of change and goals of the project were determined at the end of both meetings. Mainly, the central issue for the analysis that was agreed by both stakeholders was to get insights in the variation of the healthcare processes in the gynecological-oncology department. The aim was to visualize and analyze the variation in order to reduce difference in patient outcomes. For that reason, the actors stated four topics as the focus of the analysis. These topics were recognized as the priorities of the analysis, and some of them, as the KPIs that had to be answered at the end of the analysis. The proposed topics were:

- Level of standardization in the clinical pathways
- The collaboration between departments
- Logistic insights (like the time in process and resource utilization)
- The compliance of a policy in the department which states that the patients have to be seen always by the same doctor

From this list it was possible to establish the first set of KPIs which were: the average time in process of the patients, the resource utilization and the number of doctors seen by a patient. The details about the comparison between the activities proposed in the method and the activities performed at the AMC in this phase are shown in APPENDIX C Table 13.

The only proposed activity in the method that was not performed during this phase was the activity of obtaining information about the healthcare program. This was due to different factors. The main reason was that these sessions are intended to present the benefits of this project to the healthcare actors. These stakeholders have usually limited time due to their responsibilities in the hospital, and getting the general knowledge from a healthcare program could take a lot of time if the analysts are not familiar with the medical procedures. However, this preliminary information could save a lot of time and prevent mistakes during the analysis. Therefore, this activity was carried out after the introduction session phase and during the analysis phases. The suggestion in the method is to create a new phase for getting this information after the introduction session but before the preliminary analysis phase. Additionally, as a general recommendation for future projects, the introduction session should be performed, if possible, only once and including all the different stakeholders to generate the best possible discussion and feedback about the project. The following section will present the first set of results of the preliminary analysis. The topics proposed by the medical specialists during the introduction session were investigated in this phase.

3.3 Preliminary analysis

The obtained results in this preliminary analysis phase are linked with one of the four topics proposed by the AMC for the reason that the AMC was looking for process information in these areas. Before the analysis, certain level of aggregation was added to the data. The aggregation is an activity suggested as part of the pre-processing activities during the build database phase of the method. However, this activity was identified as necessary until this phase. The strategy for the aggregation process was agreed with the AMC to be as follows: to rename the activities and to group events at the level of a visit to a certain department per day. In other words, only one activity with the same new rename was kept per day for the same patient. This rename was done mainly by changing the name of the activity into the English name of the department that did the activity but also grouping some activities. APPENDIX D shows how this renaming was done listing the original name of the activities and the new one. This aggregation strategy was followed to clearly see if there were patterns in the data. The justification for using the aggregation was to obtain results and process knowledge at the department level, and maybe, from there, look for more specific visualizations just in certain parts of the process.

The method proposes in this phase some predefined visualizations and analyses using the tools. However, the visualizations were not obtained automatically from the tool. Therefore, the usage of the MagnaView tool had to be learned in order to produce these visualizations. It is important to mention that a few visualizations were not developed during this phase as the method proposes. At the end of this section some paragraphs are dedicated to explain why some views were not produced in this validation process as the method proposes.

This section only presents the views and results produced during the analysis. The feedback and discussion of this analysis phase is discussed in the next section which covers the presentation of the preliminary results to the medical specialists in the preliminary meeting phase of the method.

Figure 4: All activities visualization - Preliminary analysis

The first visualization produced presents the overview of all the activities for the treatment process as generated by the MagnaView tool. Figure 4 shows this visualization. This picture illustrates that "Nursing Ward H5Z", "OC Gyn Onc" and "Lab" activities are the most common activities among the patients in

the healthcare process because of their size in the figure. Additionally, each activity presents, at the bottom of it, its percentage from the total of activities performed in this dataset.

The first and last activities visualizations are illustrated in Figure 5 and Figure 6 respectively. These pictures provide information on how patients enter and leave their treatment process.

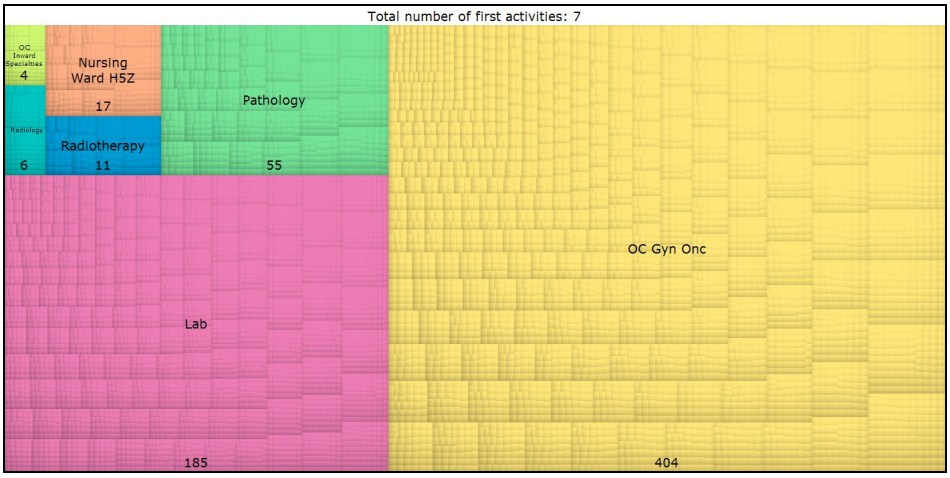


Figure 5: First activities visualization - Preliminary analysis

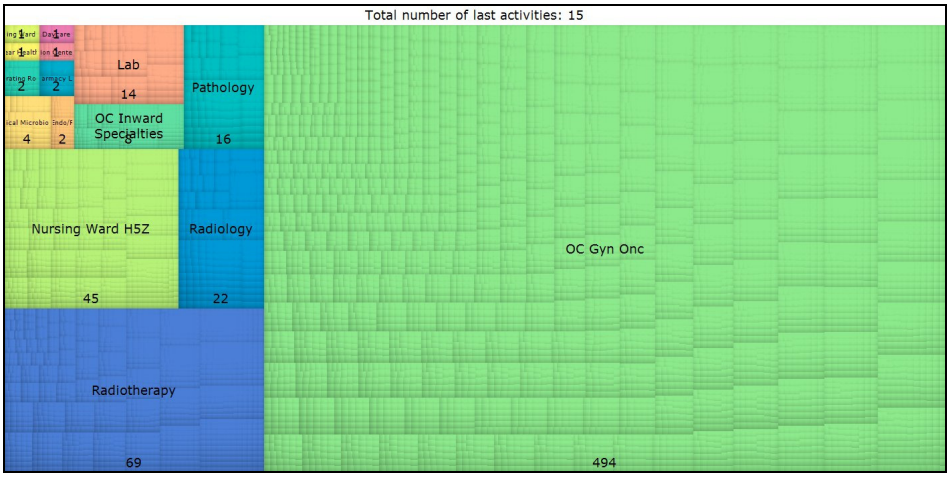


Figure 6: Last activities visualization - Preliminary analysis

As is shown in Figure 5 there are only seven first activities in the aggregated dataset. The number at the bottom of each colored square represents the number of patients that initiate with that specific activity. According to some feedback from the medical specialists during this preliminary analysis phase, the patients that start with a “Nursing Ward H5Z”, “Radiotherapy”, “Radiology” or “OC Inward Specialties” activity were mainly patients that are in the dataset but not at the beginning of their process, meaning that their first activity was not inside of the analyzed timeframe of the dataset. This fact could be used in further analysis if the intentions are to have more detail on this matter. Furthermore, in Figure 5 and Figure 6, the users can select the initial or final activities that are of interest for the analysis. These filtering options were added to the MagnaView file so the users can have an interactive utilization of the tool during the meeting in the next phase.

Figure 7 shows the view of the aggregated activities ordered in a chronological way per patient. This picture presents, in the horizontal axis, the activities performed for the patients during their process in the hospital of the whole dataset. The intention of this view is to visualize common patterns among patient processes and to observe the level of variation in the treatment. As is shown in Figure 7, it is difficult to identify clear patterns of activities. However, for the patients with more than ten activities in the process, it is possible to observe a certain level of standardization. Approximately around the third and ninth activity for this group of patients, it is possible to observe almost a vertical line of blue activities. These blue activities represent first the "Radiology" and then the "Operating rooms" activities. Then the dominant color in the picture is the green which stands for the "Nursing Ward H5Z" activities due to the fact that the patients are recovering from surgery and treatments. As far as the variation is concerned, the patients with less than ten activities have a steep slope while the rest of patients present a rather smooth slope, increasing as it moves down into the figure. This means that patients in the bottom part of the figure have more complex treatment processes, thus more activities and variation.

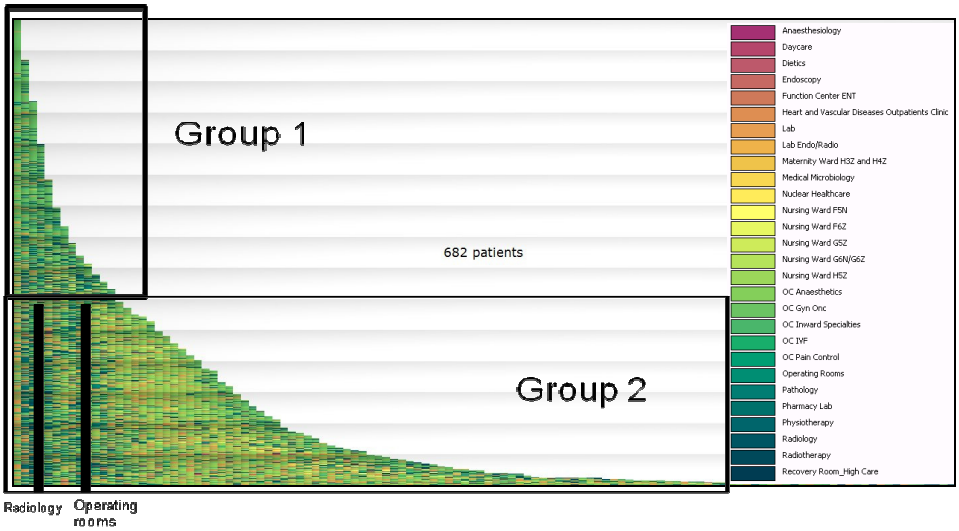


Figure 7: Patterns (sequential) - Preliminary analysis

Figure 7 was also extended with some extra analyses. Some patients in this figure had activities for more than one year. Sometimes these activities were only follow-up activities for one specific treatment. Therefore, the process-innovation manager suggested reducing the analyzed time for this view. As an example, he suggested to create one view including only the first 21 days of treatment for each patient. The option to change the length of days was added to the MagnaView file so the medical specialists can change it if they think that a different duration would be more interesting to analyze. Moreover, this type of analysis was also used during the 2nd analysis phase as it further explained in the following sections.

Figure 8 shows the activities grouped together according to the type of activity. In this picture it is possible to visualize again the complete dataset. Additionally, we can easily detect that the brown activities (Lab activities) and the green activities (Nursing Ward H5Z activities) increase as the number of activities per patient also increases. This behavior can be observed for patients with more than ten activities in their process. For the patients with less than 10 activities, the predominant color is another tint of green related with the OC Gyn Onc activities (appointments with the doctors).

The visualizations presented so far are proposed in the method as important views for getting process information. Furthermore, these views were included as part of the information shown to the medical specialists that tries to answer the first two topics described in Section 3.2. These topics are the level of standardization in the clinical pathways and the collaboration between departments. These visualizations help in getting insights of the current variation in these two general topics.

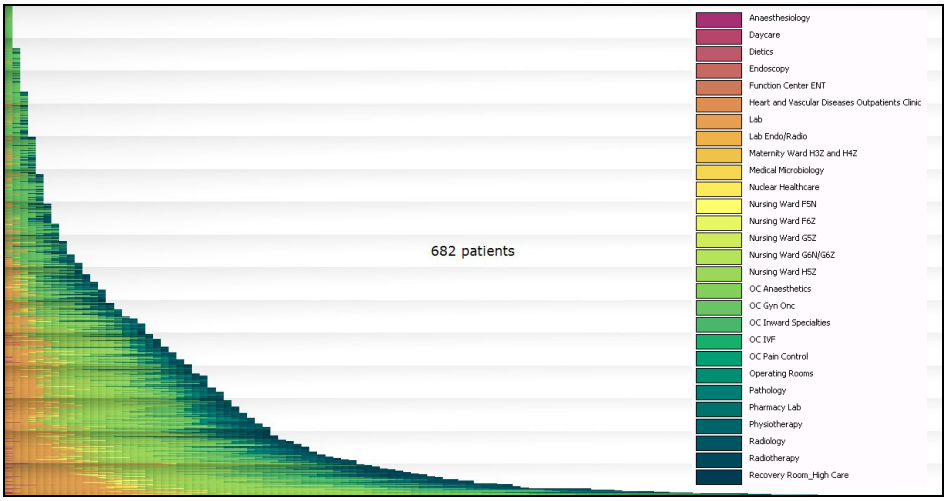


Figure 8: Patterns (grouped) – Preliminary analysis

The next step in the analysis was performed using the process mining tool. Process mining is fully described in Section 2.1. According to the method for healthcare process analysis, the ProM tool can be used to get more process insights of the treatment process. In the method, this goal is achieved by producing process diagrams of the data. To do so, the method proposes the usage of the following plug-ins in a sequential order:

- First, the *Self-Organizing Map* (Song et al. 2008) plug-in which cluster patients with similar patterns.
- Second, the *Performance sequence diagram analysis* (Hornix 2007) to check the number of patterns in the process for the groups clustered by the previous plug-in.
- And finally, the *Heuristic miner* (Weijters et al. 2006) which is a tool that mine processes producing process models. These models were also produced for the groups of patients defined by the clustering plug-in.

For that reason, those were the analysis performed with the process mining tool in the validation of this phase.

First, the possible groups of patients were investigated using the *Self-Organizing Map* plug-in. This analysis was performed before obtaining any process model and pattern information because this clustering tool can group patients with similar patterns and maybe resulting in more understandable models. According to the results of this clustering tool, the patients were divided into two groups: one of 69 patients and the other of 613 patients. The resulting visualization of this clustering plug-in is shown in APPENDIX E.

The second part of this analysis was done using the *Performance sequence diagram analysis* plug-in. First, according to the results of the largest patient group of the SOM plug-in (613 patients), there are 306 different patterns in the 613 analyzed patient processes. The results also show that 44,2% of patients had a unique pattern. The most repeated pattern had a frequency of 157 times but it was including the patients that only have an "OC Gyn Onc" activity in their process. Second, the smallest group of patients was investigated (69 patients). In this group, the results presented 49 different patterns and 63,8% of unique patterns. The most frequent pattern was repeated 15 times and was again for patients that only have an "OC Gyn Onc" activity in their process. Therefore, a high level of variation was found even in this aggregated dataset.

As was mentioned before, the purpose of the process mining results in the method is to gain more process insights by means of process models produced by the ProM tool. Therefore, the final step was to analyze the process models for both groups of patients extracted by the *Heuristic miner* plug-in. The resulted process models presented difficulties to fully understand them due to the fact that they contain a lot of activities and arrows in different directions. In other words, these models are containing a lot of variability in the processes. To see the process models in detail, go to APPENDIX F.

The remaining two topics pending for analyses so far are the logistic insights and the information about the number of different doctors seen by the patients during their diagnosis and treatment processes. These two topics were included in the last step of this analysis phase. Consequently, the calculated KPIs were:

- The average time in process for patients
- The resource utilization
- The number of doctors seen by each patient

The average time in process was simply calculated by the difference in days between the first and last registered date for each patient. The results are presented in Figure 9.

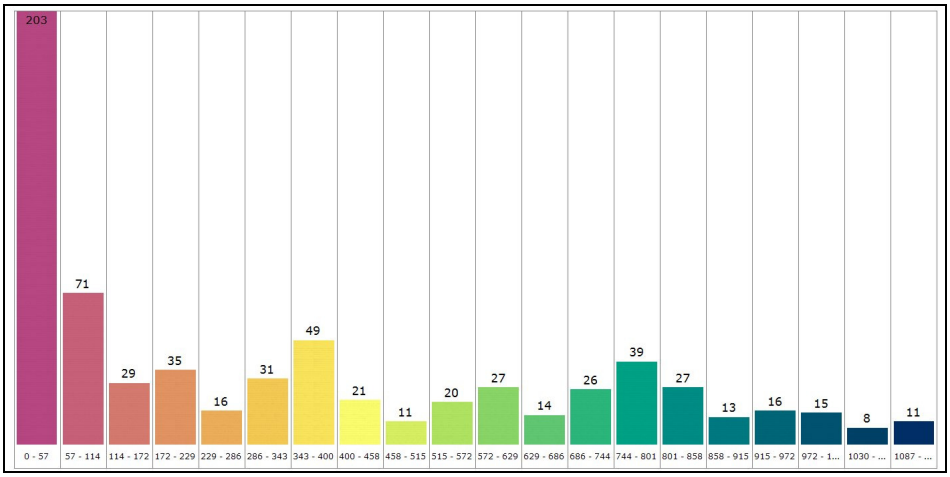


Figure 9: Average time in process for patients (days)

In the bottom part of the graph it is possible to see a period of approximately 57 days and the number in each bar represent the number of patients that are inside the period defined in the bottom. For example,

the first bar shows that 203 patients have an average time in the process between 1 and 57 days. This period of 57 days was randomly assigned just to show the results in bar-chart visualization.

The second KPI was the resource utilization. This indicator was calculated using the timestamps available in the dataset. The resource utilization was defined as the percentage, from the total days in the timeframe, in which each department had at least one patient per day. It is important to mention that these utilization percentages were only for the patients of the gynecological-oncology department. For instance, the results show that 91% of the analyzed time, the "Nursing Ward H5Z" department cares for at least one patient.

In order to calculate the last KPI a different file was required from the AMC. This file contains the information on the appointments to the hospital for each of the analyzed patients, like the date, length, name of the doctor and type of appointment. This file was also loaded into the MagnaView tool. The results show that 315 patients were only seen by one doctor while 3 patients saw 7 different doctors. For details on the visualizations of the second and third KPIs see APPENDIX G.

The results of the KPIs analyzed during this phase were validated by means of case analyses. Therefore, for each KPI, a number of specific patients were analyzed individually in order to corroborate that the data was correctly calculated in particular cases.

The next visualizations are not included in the method; still they were developed in this phase because it was thought that they can contribute in gaining useful process information. Figure 10 shows the information on the activities per patient but this time each column represents a day in the hospital. Therefore, the size of the colored squares depends on the number of activities performed for each patient on the same date. By means of this figure it is possible to see that in the second date of the patients with more activities, there is a blue line. This blue line is mainly Pathology and Radiology activities. Thus, these activities are mainly performed on the second date of patients. This figure contributes to get insights in the topic of the standardization in the clinical pathways. However, it also helps in seeing the daily patterns related with the collaboration between departments for the same patients.

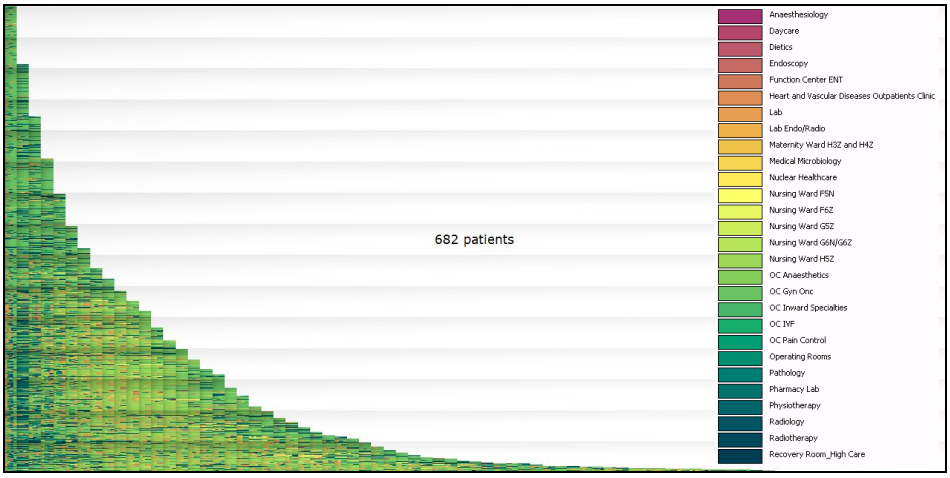


Figure 10: Activities per day per patient - Preliminary analysis

Figure 11 was created using the same strategy of Figure 10 but in a different presentation in order to identify patterns per date of patient and, consequently, the collaboration between departments per date of patient. Thus, Figure 11 tries to show and identify the most common patterns in the dataset. By means of this view, it was discovered that there are 403 patients that have at least one date with two activities, 362 patients have the combination of Lab and OC Gyn Onc activities which was the most common collaboration between departments in one date of patients. This combination can be recognized in Figure 11 where, inside of the same small squares which represent the patients, it is possible to see a brown and a green activity in the same column. The filter option to choose which departments to analyze was added in this view.

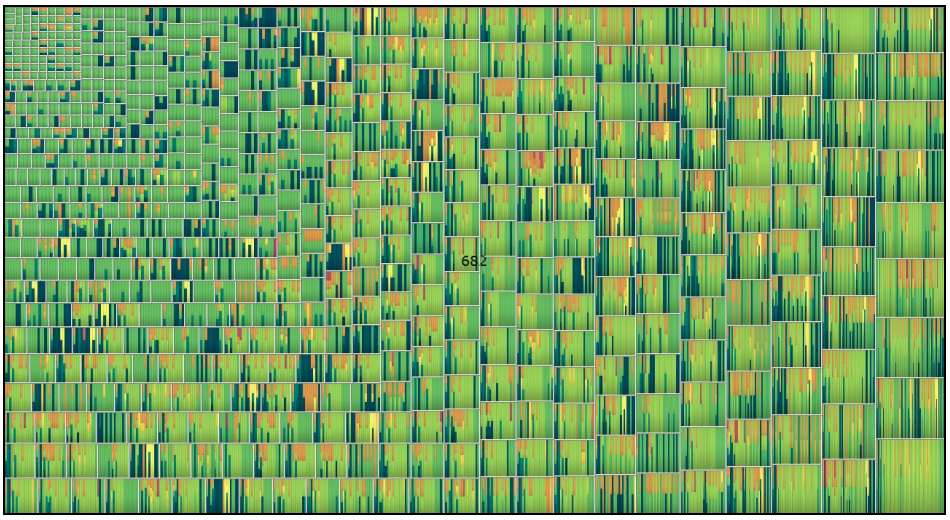


Figure 11: Patterns per day - Preliminary analysis

In order to continue with the investigation in the collaboration between departments, Figure 12 was also elaborated. This figure shows the distribution of the handover of work. The explanation of the figure is that the activities in the title handover work to the colored activities in its square. Additionally, the size of the colored squares is related with the frequency of the handover.

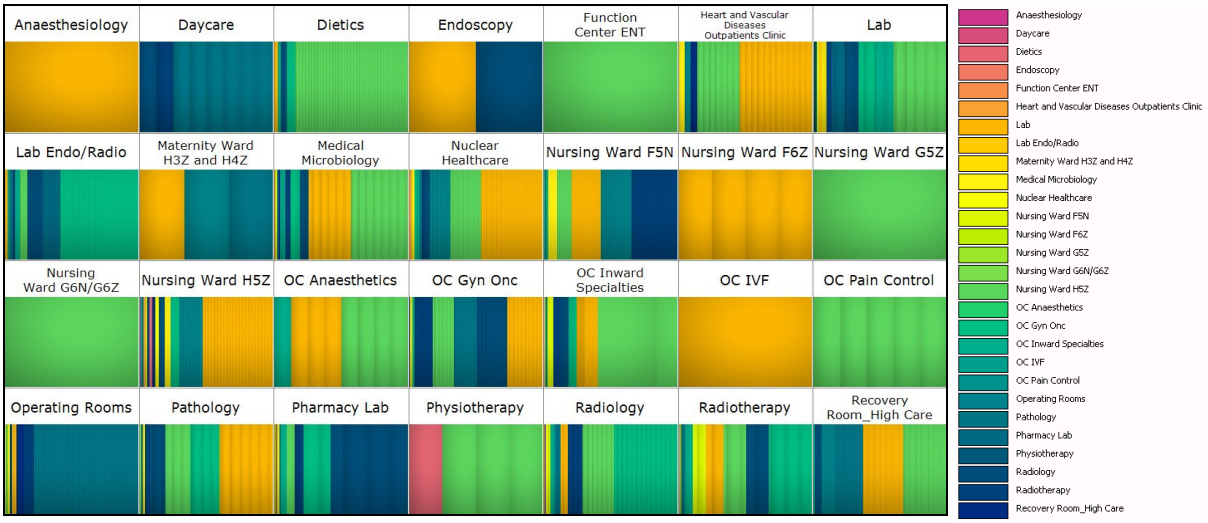


Figure 12: Handover of work - Preliminary analysis

According to the method all the analyses are done with MagnaView except for the process mining results which uses the ProM tool. The details in the comparison between the analyses suggested for this phase in the method versus the analyses elaborated in this validation project can be seen in APPENDIX C Table 14.

The important issues that need an extra explanation in this phase are:

- The first five steps of the method include the following views: "All activities" (Figure 4), "First activities" (Figure 5), "Final activities" (Figure 6), "Patterns sequential" (Figure 7) and "Patterns grouped" (Figure 8). These visualizations were developed as suggested in the method. Both patterns views, sequential and grouped, were extended with extra analyses added in the views. These additions were on adding options to limit the analyzed time on the data (e.g. only certain period of time for each patient) and on filtering activities out of the analyses.
- According to Riemers (2009), during the development of the method for healthcare analysis, the combination of the mentioned plug-ins of the ProM tool used to get process mining results added some important process insights to the medical actors. However, during this preliminary analysis phase of the validation process, the resulted process models were too complex to understand and to made conclusions from them. Thus, these analyses were repeated during the 2nd analysis phase of the method with a different dataset in order to produce more understandable process mining results. The changes made to the dataset and the new analyses with the ProM tool are fully explained in the following sections.
- The method proposes four different views that analyze the processes before and after certain centered activity. These visualizations were not developed during this phase due to the fact that no centered activity was identified by the medical specialists as a relevant activity for this type of analysis. During the preliminary meeting, the specialists identify an activity for this steps and the centered view was developed in the 2nd analysis phase of this validation process.
- For the analyzed KPIs, the method suggests that each KPI must explain and describe the general information of it, the overview, results and validation process. These four points mentioned in the KPI's analyses were also followed during this validation process. Each KPI has explained its meaning, how it was calculated, the results and its validation.
- The extra visualizations and analyses performed in this phase ("Activities per day per patient" view - Figure 10, "Patterns per day" view - Figure 11 and "Handover of work" view - Figure 12) were presented to the different stakeholders in the next phase in order to determine if they provide useful process information so they can be added to the method.

All these results were calculated during the preliminary analysis phase of the method. Once the results were ready, the preliminary meeting was planned and scheduled. This meeting is aimed to give a small presentation to the actors in order to show them what the tool can do and to receive feedback on which specific areas are the most interesting in going deeper in the analysis. Next section describes all the details regarding this meeting.

3.4 Preliminary meeting

According to the method, the goals of this meeting are to:

- Explain and discuss the initial results

- Use the feedback to improve the analysis
- Let the actors use the tool and see its benefits
- Determine which activities should be kept or removed from the analysis
- Determine follow up steps and focus process analyses

These goals are explained individually in the following paragraphs. The method suggests that the results should be presented to 3 actors so the goals of the meeting can be reached without generating too much discussion. The results of the preliminary analysis were presented again in two different meetings. The first meeting was with the leading manager of the process innovation department who checked the obtained results. His feedback was focused in how to present the results to the medical specialists in the coming meeting and not in determining follow-up activities for the 2nd analysis phase of the validation process. The second meeting was held with the leading medical specialist of the department and two PhD students involved in the gynecological oncology department. This preliminary meeting has the objective to present the results in order to get feedback on the views and analyses developed so far, and more specific questions to continue with the analysis. Thus all the views presented in the last section were explained during these meetings.

After the presentation, the interaction with the tool was limited due to the fact that the available time for the meeting was relatively short. However, the actors saw the usefulness and the interactive characteristics of the tool, especially when filtering events and analyzing the resulted visualizations.

In determining which activities should be further analyzed or removed from the analysis, the next actions were recommended. First of all, the small group of medical specialists agreed that the aggregation of data was good to have a general idea of the clinical pathways. However, it did not show a good level of detail from which it was possible to obtain interesting insights. Thus, they proposed to work with a different level of aggregation. Basically, the proposal was to work with the original dataset but filtering out the administrative activities and joining the lab activities performed in one day for the same patient as only one activity. They justify this joining of activities with the statement that a lot of small lab-tests can be performed in one day for one patient but all of them correspond to one lab activity for the patients. Furthermore, the analysis shown in Figure 7 (Patterns sequential view) was interesting but, according to them, the scenario was not well delimited. First, only the patients that are since the beginning of their process should be included. In other words, the patients that are in the middle or at the end of their clinical pathways between the dates that consider the dataset must not be shown. Second, the analyzed activities that should be visualized must be only the activities performed in the first three months of the patient treatment processes.

The feedback received from these visualizations can be summarized in the following points:

- The all, first, and last activities visualizations (Figure 4, Figure 5 and Figure 6) should be shown only for the patients that were at the beginning of the treatment process.
- The patterns views (sequential and grouped, Figure 7 and Figure 8 respectively) should be shown including the first 3 months of treatment for patients that were at the beginning of the process.
- The process mining results (APPENDIX E and APPENDIX F) were not shown in this meeting due to the fact that no additional process insights were identified from them. Further analysis was conducted in the 2nd analysis phase to find better results.

- The surgeries were highlighted as a good candidate to use in the centered activity views that were not yet performed in this analysis.
- The KPIs of the average time in process and the resource utilization (Figure 9 and Figure 36) were not useful, so the medical specialists proposed for the next analysis phase to develop the visualizations of the average time for surgery and the average time from surgery to radiotherapy in substitution of the last KPIs.
- The "doctors seen by patient" visualization (Figure 37) was really interesting but it must include only the patients and appointments of the gynecological oncology department. Therefore, this view has to be also modified.
- The feedback on the extra analyses was: The "activities per day" view (Figure 10) was also useful and used in further analysis; The "patterns per day" (Figure 11) was difficult to explain and understand but the concept was good to see the patterns in the processes; Finally, the handover of work (Figure 12) was also interesting but the view contained too much information. The doctors ask for this view to only show the most important activities.

Based on the previous feedback, the medical specialists that were present during the second meeting of this phase (the leading doctor of the department and two PhD students) mentioned some specific questions for each of the topics presented in Section 3.2 as the follow-up activities and focus process analyses to be investigated in the 2nd analysis phase of the project. Additionally to these questions, the specialists asked, wherever was possible, to compare the results mainly of three DBC-codes: M11, M13 and M16. To see all the specific questions elaborated during this meeting see APPENDIX H.

From the steps proposed in the method, only one activity was not completely performed in this phase during this validation process. This activity was letting the actors to "play" and use the tool, mainly because of the limited time available for these meetings. However, it is important to keep this activity in the method because it can help the healthcare actors to better understand the tool and to go deeper in the focus analyses. APPENDIX C Table 15 presents the comparison between all the activities proposed in the method for the preliminary meeting and the activities performed during the validation project at the AMC. Additionally, it is important to mention that the follow up analyses established during this phase of the method are critical to achieve good analysis results at the end of the project because this follow up activities are used to perform the analyses on the 2nd analysis phase and the results of that phase are the final results of the project. In the case of this validation project, the process manager did not provide with feedback that could be use in follow up activities during the 2nd analysis phase due to the fact that his feedback was intended to improve the way of presenting the results to the medical specialists and not in suggesting how to continue with the analysis in the 2nd analysis phase of the project. Moreover, the decision of not including more doctors of the gynecological oncology department in this preliminary meeting, who are the main responsible of the activities performed to the patients in this department, could have a negative impact in the involvement of these stakeholders in the remaining stages of this project and to perform analyses that are not real issues for the medical actors. Therefore, the method should make a stronger recommendation to include managers and doctors in only one preliminary meeting in order to determine, through their feedback, how to improve the analysis. Next section presents the results obtained during the second analysis phase of this graduation project.

3.5 2nd analysis

The 2nd analysis phase focuses on the feedback from the preliminary meeting. It is important to mention that all the views and analyses for the specific questions mentioned as the follow-up activities in the preliminary meeting feedback (APPENDIX H) were produced. All the results and analyses were included in the final documentation elaborated for the hospital, as explained in Section 3.7. However, in this section the focus is on the views considered as the most important ones. These most important views were selected because all the results were calculated from them. In other words, the general lay-out of a few visualizations was used more than once but to filter different activities or to compare different results in order to answer the specific questions. This section only shows the views once in order to continue with the validation process of the tool.

Before starting with the 2nd analysis, the aggregation of the lab-activities was done as suggested by the medical specialists in the previous phase. Therefore, some pictures have relatively changed as the ones presented in the preliminary analysis phase. The new used dataset contained fewer cases than the previous one and it was loaded again into the tools for the analysis. Additionally, the following analyses were done only to the group of patients that starts their process in the timeframe of the dataset, as recommended by the medical specialists during the preliminary meeting. This new group includes only 362 patients. Additionally, the coloring and analysis continue at the department level. So, in the following views it is possible to see all the activities performed but the analysis is not on the individual activity but on the department that produces it.

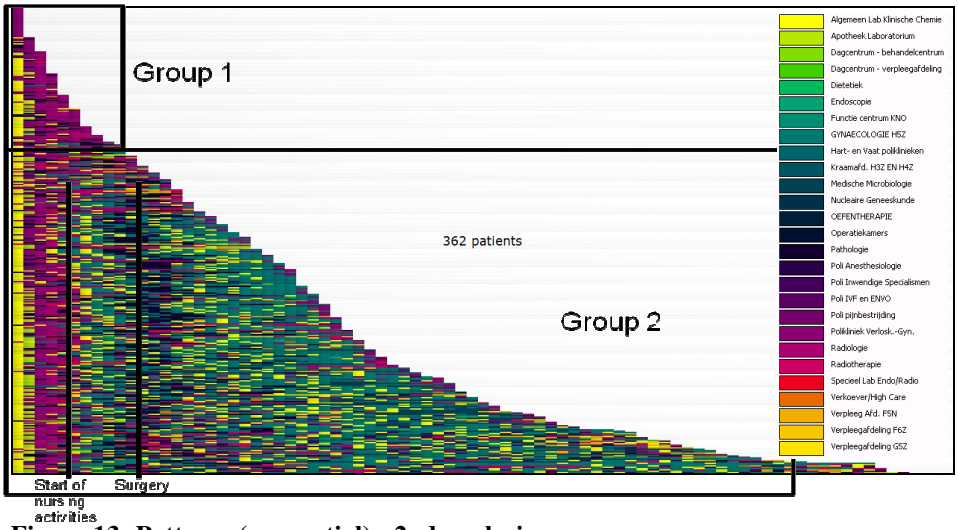


Figure 13: Patterns (sequential) - 2nd analysis

Figure 13 shows the "patterns sequential" view but only for the patients that were since the beginning of their process. This figure presents a clearer level of standardization compared with the same figure produced in the preliminary analysis phase. In Figure 13, it is possible to recognize two different groups of patients. The first group would be the top-left-corner which presents patients with mostly pink activities. These pink activities are the outpatient activities, the appointments with the doctors. This group represents the patients that are not/yet not being hospitalized and because of that are the patients with less reported activities in the process. On the other hand, the rest of the patients can be joined in a different

group. These patients are the ones with more activities in the process. They also present, at the beginning of their process, appointments with the doctors and radiology activities. However, after certain meetings with the specialists these patients start with the nursing activities leading to the blue activities which can be recognized almost as a straight vertical line in Figure 13. These blue activities are the surgery activities. After the surgeries, it becomes more difficult to find standardized activities maybe due to the fact that the recovery of patients depends mainly on each individual condition or because we are only seeing the first three months of treatment. Additionally, we can see, in both groups of patients that almost all of them have as their first activity a yellow activity. These yellow activities are the lab activities. It is obvious that doctors need some lab-test results in order to establish the real condition of the patient to start with the treatment.

Figure 14 presents the most frequent departments for patients that are in DBC's M11, M13 and M16. As is shown in Figure 14, the data can be divided first according to patient's DBC, and then by the department in which each activity was performed in order to facilitate the comparison between data. Therefore, we can see that "Gynaecologie H5Z", "Polikliniek Verlosk.-Gyn.", "Algemeen Lab Klinische Chemie", "Radiologie", "Pathologie", "Medische Microbiologie", "Radiotherapie" and "Operatiekamers" are the most frequent departments for patients in these DBCs.

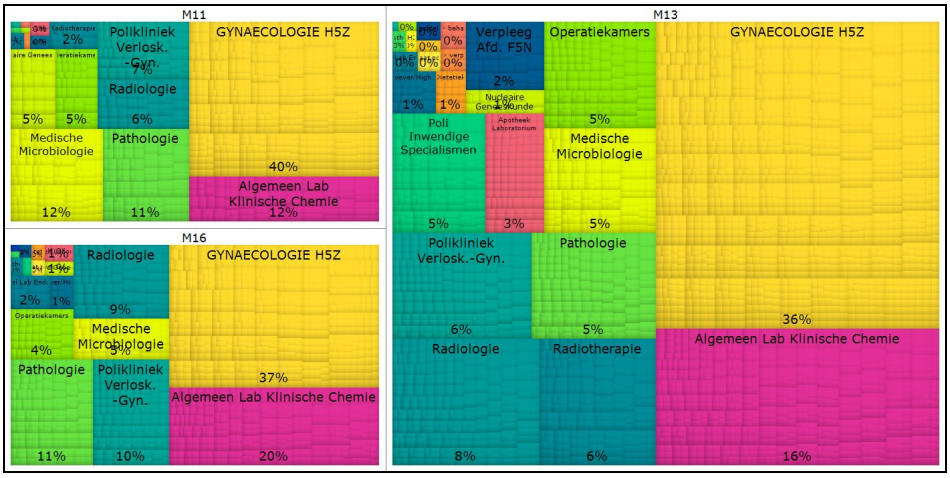


Figure 14: All activities - 2nd analysis

The first and last activities visualizations were not required by the medical specialists to be repeated in this analysis phase due to different reasons: First, the "first activities" view was not needed due to the fact that the medical specialists use it in the preliminary meeting session to identify the patients that were since the beginning of their process. However, these patients were easily identified with the file that contains the appointment with the doctors (one type of appointment referred to the first appointment with the physician). Still, this view was elaborated to check the results. The new dataset contained only 5 patients out of 362 that have their first activity not in the "OC Gyn Onc", "Lab" and "Pathology" departments. These patients were kept in the analysis because they could represent some exceptions due to the fact that they have their first appointment with the doctors included in the analyzed dataset. Second, the "last activities" view was not necessary because the analyzed activities in the patients changed to only their first three months of treatment. This change made this specific analysis useless.

Some questions were generated during the preliminary meeting about the frequency of surgeries and radiotherapies. Figure 15 shows the results of the most common radiotherapies found in patients with the DBC-codes of M11, M13 and M16. As is shown, each DBC-code presents a histogram which calculates the frequency of the radiotherapies found among its patients. For example, it can be seen that the most common radiotherapy in the patients of M13 is "Teletherapie – megavolt fotonenbestralingszitting".

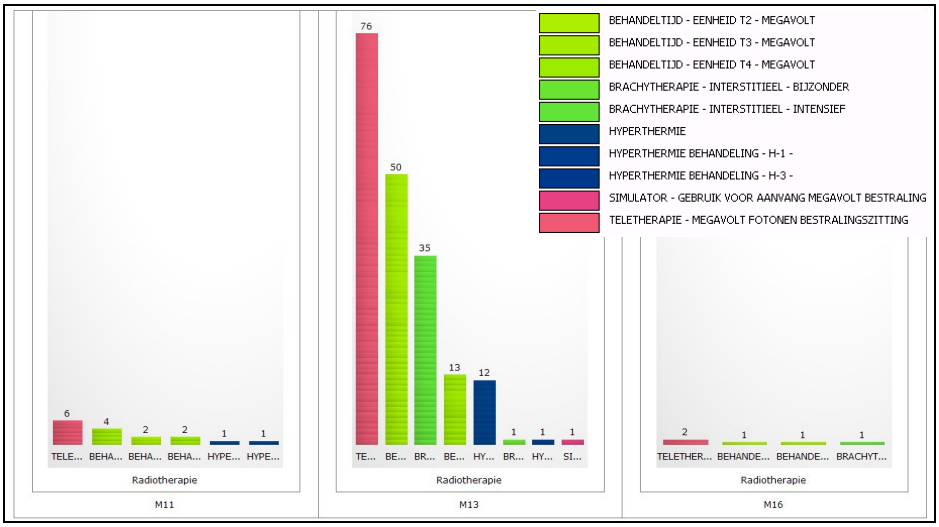


Figure 15: Most frequent activities (radiotherapies) - 2nd analysis

In this 2nd analysis phase, the medical specialists pointed out the surgery activities as an important activity that could be used for the "Centered activity" views. However, from these types of views proposed in the method, the Centered activity with patterns (sequential) view was the only one produced due to the fact that the hospital was only interested in the current order of all the activities before and after a surgery.

As became clear from Figure 13, it is possible to find some level of standardization in the group of patients analyzed in this project. Figure 13 showed that, for the group of patients with more activities in the process, the surgeries can be the limit of some kind of standardization in the clinical pathways. Before surgery, patients have lab tests, appointments with the doctors and some nursing ward activities, all of them in more or less a standard way. After the surgery, it becomes more difficult to establish standardization in that picture. Figure 16 shows with more detail the activities performed before and after the most common surgery in the dataset.

Thus, Figure 16 illustrates the activities that were performed before and after the surgery "Vagina – Toucher onder anesthesie" (the most common surgery in the dataset). This surgery is represented by the centered blue rectangle and, in a horizontal perspective we can find the activities performed for each patient before and after the surgery in a chronological way. As is shown in this figure, the standardization found in Figure 13 is confirmed here. Before the surgery takes place, there is a clear standardization of activities. Mainly, the order of activities before the surgery is as follows: lab tests, outpatient activities, radiology, nursing ward and then, the surgery. However, it is also recognized that the process previous to this surgery is not completely standardized because it has some variation easily observed in the figure. On the other hand, the activities after the surgery are showing a lot of variation and no easily observed pattern of standardization.

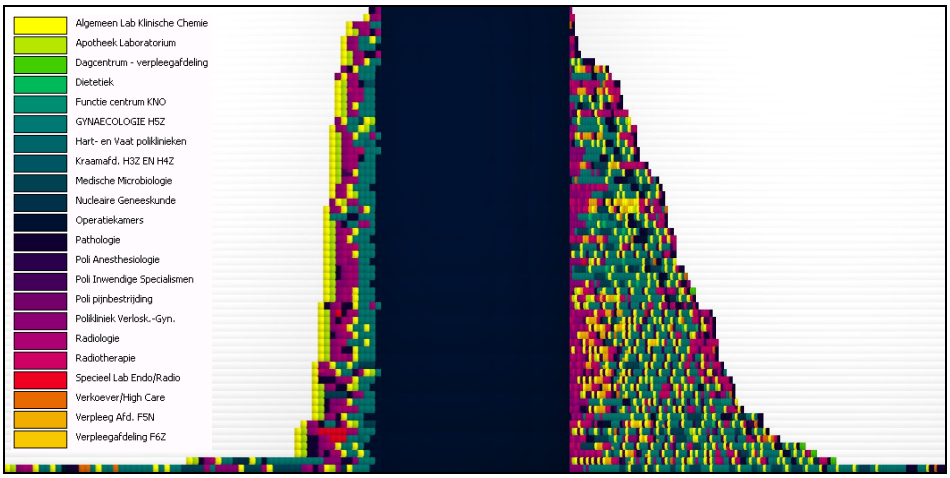


Figure 16: Centered activity with patterns (sequential) - 2nd analysis

The process mining results obtained in this 2nd analysis phase came from using the same plug-ins as in the preliminary analysis phase but for the new dataset. The reason to use them again was to verify if the results from this new dataset can contribute to get useful process insights and to answer one specific question related with the patterns in the data. Thus, the order of plug-ins was the same as in the preliminary analysis phase. First, the clustering results (*SOM* plug-in) divided the patients into two groups, one with 315 patients and the second one with 45 patients. For more details on the obtained visualization after the clustering process see APPENDIX I. Second, the next step was to calculate the patterns in these two groups of patients. First, according to the results of the *Performance sequence diagram analysis* plug-in, the largest group with 315 patients had 206 different patterns and 47,5% of the patients had an unique pattern. On the other group of patients, 44 patterns were found for the 47 patients included in this group, resulting in 87% of unique patterns. Finally, the *Heuristic miner* plug-in was used in both groups of patients to obtain the process diagram. For the details in these process diagrams see APPENDIX J. Even though the produced models in this phase are simpler than the produced during the preliminary analysis phase, the results did not show clear process information for medical stakeholders. These plug-ins add important additional process information during the project that developed this method, however in this validation project they do not contribute to increment the understanding of the processes among medical stakeholders. Therefore, it seems that additional research is needed on the ProM tool in order to determine, in a more general way, which analysis can be carried out with this process mining tool to obtain useful healthcare process information. These issues are investigated in Chapter 5.

Figure 17 presents the visualization of handover of work for the most frequent departments. In this picture, the data is divided in the three most important DBC codes. Then, each DBC-code has the most important departments, and the handover of work is shown in the colored area below each department. For example, in M13 “Radiologie” handovers work mainly to the yellow department, and some to the purple and green departments.

As is shown in Figure 17, the most frequent departments make their handover of work only to five departments (“Algemeen Lab Klinisch Chemie”, “Pathologie”, “Poli Inwendige Specialismen”, “Polikliniek Verlosk.-Gyn” and “Radiotherapie”). Moreover, in the DBC-codes M11 and M16 the handover of work is only done to three different departments (“Algemeen Lab Klinisch Chemie”,

"Pathologie", and "Polikliniek Verlosk.-Gyn"). Additionally, it can be also recognized that around 80% of the handover of work made by the most frequent departments is done to the lab-department who mainly handover work to the "Polikliniek Verlosk.-Gyn".

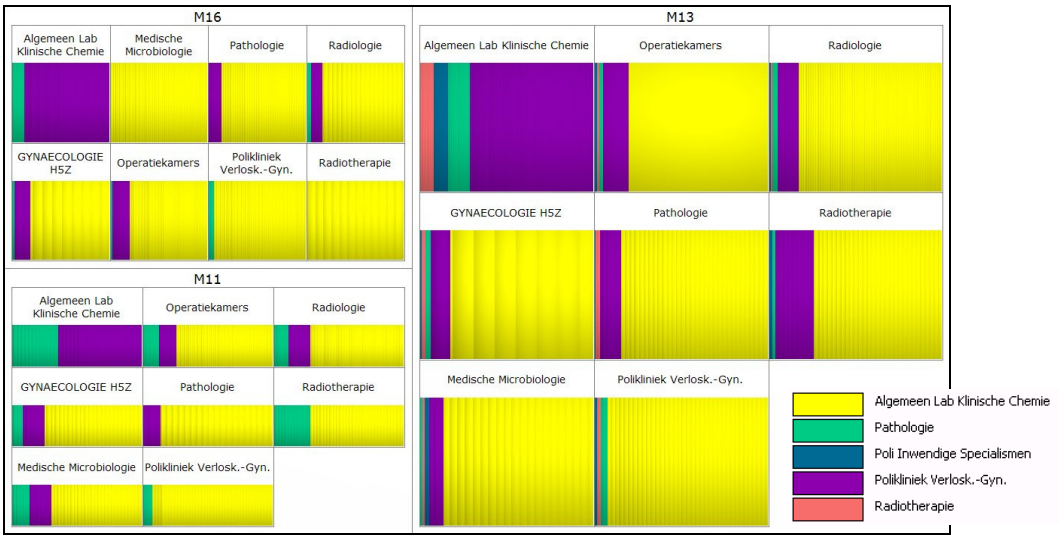


Figure 17: Handover - 2nd analysis

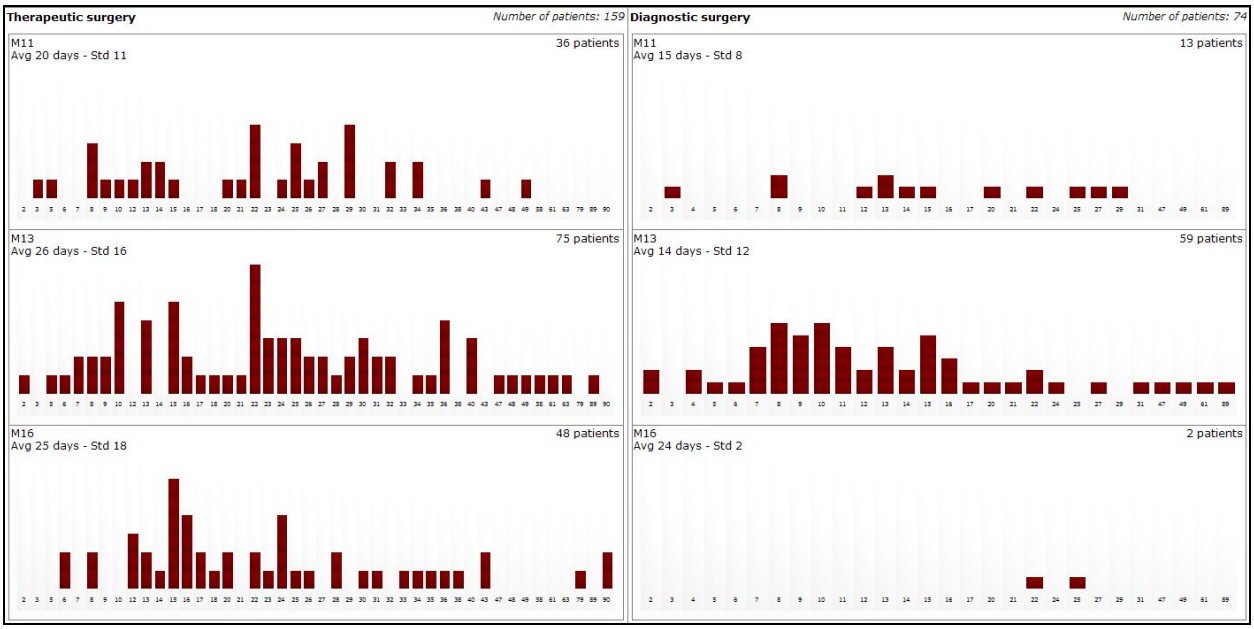


Figure 18: Time for surgery per DBC-code

Figure 18 presents the first of the logistic performance indicators developed during this phase. This indicator is the average number of days that patients had from their first appointment with the doctors in the hospital to their first surgery intervention. In this view, the data only contains the patients that had a surgery activity during their process at the AMC. As is shown, the patients are also divided according to the type of surgery performed. Two different types of surgeries were identified: the therapeutic and diagnostic types. As is shown in the figure, the patients are also divided according to their DBC-code.

Each histogram in the picture shows the number of patients in that DBC-code, the average number of days of its patients between the first appointment and the first surgery and its standard deviation. The histograms present in the horizontal axis the number of days for this performance indicator, and the height of each bar represents the occurrences of each measure. For both types of surgeries M13 is the DBC-code with the most patients. Furthermore, for the therapeutic surgeries, the average time for surgery in M13 is larger than M11 and M16. The opposite occurs in the diagnostic surgeries where M13 is the DBC-code with the shortest average time but with its standard deviation higher than the other two DBC-codes. Additionally, the average times for surgery are longer in the therapeutic surgeries for M11 and M13 than in the diagnostic type (M11 goes from 15 to 20 and M13 goes from 14 to 26 days, almost the double) while in M16 the average time for surgery is almost equal for both types of surgery, 24 and 25 days for the diagnostic and therapeutic surgeries respectively.

The "activities per day per patient" view was again used during this 2nd analysis phase. Besides the additional process insights that offers now for the first three months of treatment, this figure was also used to calculate the average time between the first and second date of patients in the three DBC-codes.

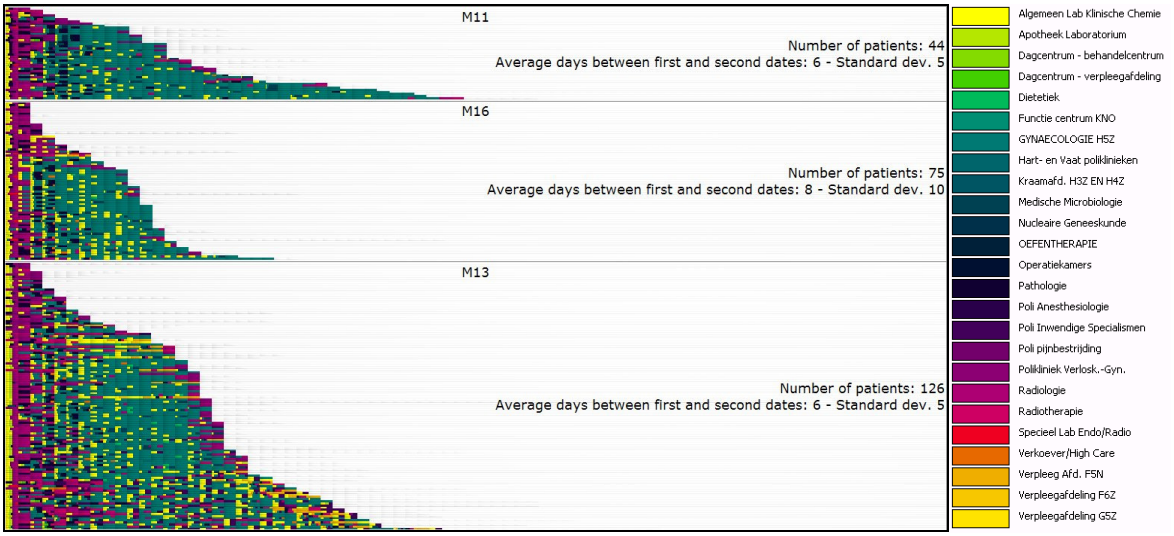


Figure 19: Activities per day per patient - 2nd analysis

Figure 19 shows the average number of days between the first and second date of patients and its standard deviation. The results show that M11 and M13 have an average of 6 days with a standard deviation of 5 days in this measure. On the contrary, M16 presents a larger average time of 8 days and also larger standard deviation of 10 days.

As was mentioned before, from this figure is also possible to establish certain level of patterns in the dates for patients. As is shown in the first column of the three DBC-codes, the main colors are the yellow and pink. From the legend of colors, these colors represent the lab and outpatient clinic activities. Therefore, patients have mainly these two activities on their first date in the process. Then, the second date has some appointments and pathology activities. This second column confirms the indication of the medical specialists about the second date which initiate the treatment base on the results from the lab-tests. From there, the nursing ward activities initiate with some surgery and pathology activities. The further dates include lab and nursing ward activities.

The analysis in the number of doctors seen by patients in the 2nd analysis phase has different results compared with the values obtained in the preliminary analysis phase. This difference is explained by the fact that the appointment dataset was linked to the raw data and some appointments were not included in the original dataset, meaning that some appointments included in the preliminary analysis were not from the gynecological-oncology department. Therefore, the results show that 190 patients saw only one doctor while 4 and 2 patients saw five and six doctors respectively. The same validation process used for the indicators in the preliminary analysis phase was implemented for the investigated performance indicators in this 2nd analysis phase. For the details on this visualization see APPENDIX K.

Figure 20 shows the activities per patient not in a chronological order but grouping the activities of the same type. This figure was utilized for different purposes during the 2nd analysis phase. It helped in analyzing the number of activities per type that were performed for different groups of patients and compare these groups that were selected according to characteristics like the number of doctors seen.

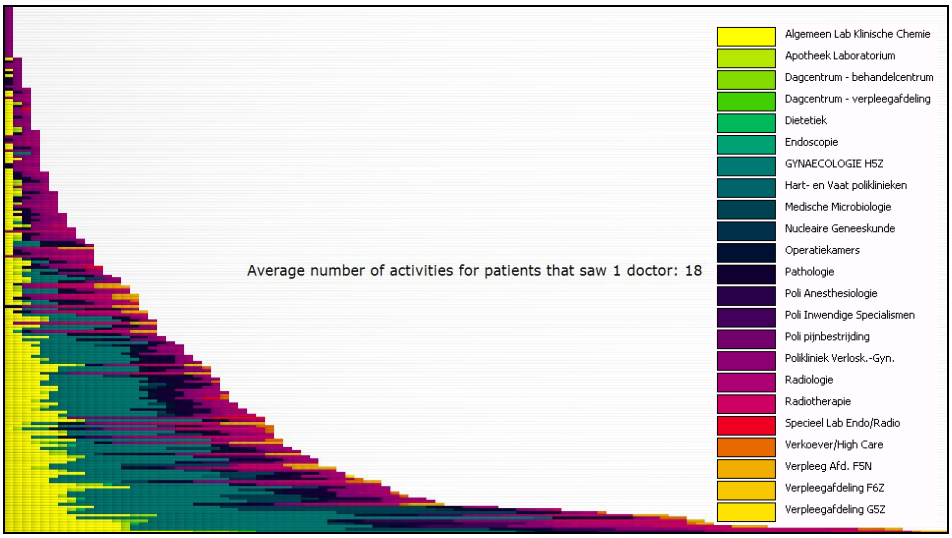


Figure 20: Patterns (grouped) - 2nd analysis phase.

Again, the method proposed the same views and analyses as in the preliminary analysis phase and the comparison between the method and the activities done during the project at the AMC can be seen in APPENDIX C Table 16.

The following list summarizes the most important differences between the method and the validation process:

- The “first and final activities” views were not produced due to the fact that they were not required during the feedback of the previous phase
- The “centered activity view with patterns (sequential)” was the only centered activity view produced during this phase because the medical specialists were only interested in finding the total and real order of activities before and after the surgeries. However, the rest of the centered activity views could be beneficial for improving or finding more interesting process information.

- In this analysis phase the extra views include: histograms to show the number and type of surgeries and radiotherapies among the DBC-codes, the handover of work of the most common departments, and the view of "activities per day per patient" to analyze certain aspects of the clinical pathways and to visualize certain average time in the dataset.

Once the analysis was done, the final meeting was planned and scheduled with a larger group of medical specialists. Next section describes the details of that meeting.

3.6 Final meeting

At the end of the analyses phases of this project, the most important results were presented to the medical stakeholders. The final meeting was again done for two different stakeholders: first, to the medical specialists and second, to some process-innovation actors in the hospital. Due to time availability, just the most important results were shown to the audience, the results that could cause a positive impact on the medical stakeholders. The steps described in the method for this phase were performed for both presentations. The details in the actions included in this final meeting are presented in APPENDIX C Table 17. The evaluation and feedback received during this final meeting are going to be fully discussed in Chapter 4 due to the fact that this evaluation and feedback are the evaluation and feedback of the complete project and that chapter includes these general conclusions of the validation process. The final step to complete the validation process was the final documentation, including all the results for the analysis, which is described in the following section.

3.7 Documentation

The documentation is the final step of the method used for healthcare process analysis. This documentation contains the description of the results and process analyses performed during the project so the actors can read it back and analyze the data by themselves. The documentation must contain two different documents. First, the project document discusses and presents all the analysis results. Second, a project file is also elaborated which contains all the data and visualizations performed. Both documents were provided to the AMC so they can have the complete documentation of this validation process and the obtained results. The final project document delivered at the AMC included the information asked by the medical specialists during the preliminary meeting. Therefore, the structure of the document was different as the one proposed in the method because of the specific process questions that it tries to answer. The comparison between the proposed activities in the method for this phase and the activities performed at the AMC are shown in APPENDIX C Table 18. The additional analyses included in the project document for the AMC that are not in the method were: the list of the most important handover of work, the list of the most common activities in certain departments (surgery and radiotherapy), and most important conclusions from the view of "activities per day per patient".

Next chapter presents the validation conclusions and remarks on the validation process performed at the AMC. Additionally, the final evaluation and feedback received from the different stakeholders on the analyses produced during this project are also discussed.

4. CONCLUSIONS ON THE VALIDATION PROCESS

As was mentioned in Section 1.3, the validation approach in this graduation project is defined as the process of checking the reliability and relevance of the method in satisfying the process information needs of the different stakeholders at the AMC. As has been shown in the previous chapter, the phases in the method proposed certain steps, analyses and tools in order to reach the specific goals of each phase. These specific analyses were performed according to the necessities of process information generated during this project at the AMC. Thus, some of the analyses were not produced due to the fact that they did not contribute to solve the process-related requirements in this project. Moreover, some extra analyses were developed in order to satisfy all the information requirements.

This chapter presents the conclusions on the validation process in order to determine if the reliability and relevance of the method was proved. First, the evaluation of the tools and views, from a technical point of view, is presented in the following section. This evaluation consists of showing which visualizations and analyses were presented as the final results of the project at the AMC, the feedback received on these views and a general analysis of the used tools, not only in the 2nd analysis phase, but during the complete project in the hospital in order to conclude if they represent a good option to include in the method. Second, an entire section is dedicated to present the opinion of the different stakeholders on the method received during the final meeting phase. This feedback is used to establish if the goal of producing results with the characteristic of being understandable was reached. This second section is also going to evaluate the steps of the method in order to determine which steps must be included in the method. Finally, some conclusions and recommendations on the method are given in the final section of this chapter. These conclusions establish if the reliability and relevance of the method were verified. Additionally, this section justifies the extra analyses presented in Chapter 5.

4.1 Evaluation of the tools and views

Due to time availability in presenting the final results to the medical specialists and to some of the process-innovation staff, only some analysis and views were selected in order to present the most important results in the final meeting. This selection was based in the information offered by each view. Both presentations at the AMC with the different stakeholders contain the same process results. It is important to mention that all the produced views were included in the final documentation delivered at the AMC. The views shown to the different stakeholders during the final meeting were:

- Patterns (sequential). This view was included because it gives a general picture of the processes in the first three months of patient treatments. From analyzing this picture it is possible to establish the level of standardization in the clinical pathways and at the same time it is possible to show the large amount of variation contained in the dataset. Additionally, this picture was used to compare the processes of different groups of patients. Mainly, to compare patients from different DBC-codes and from different ages.
- Centered activity with patterns (sequential). This view confirms the findings of the previous analysis (patterns sequential) in the standardization of activities before and after the surgery activities. Additionally, the view offers an alternative to analyze any centered activity using this format.

- Activities per day per patient. This picture is not included in the method but developed during this validation project. It was first used during the preliminary analysis phase to give more insights in the activities performed for each patient per date with the purpose to find patterns and common collaboration between departments. Some conclusions were drawn from it. However, during the 2nd analysis phase, this view was also used to calculate the time between the first and the second date of patients. Therefore, this view can be improved so it can show process information about patterns per day of patients and also waiting times between activities. These improvements are pursued in the following chapter.
- KPIs (Average time for surgery, average time from surgery to radiotherapy, average number of appointments per patient, and number of doctors seen by patients). These indicators were presented by means of histograms. In this way, it was possible to visualize the distributions in each analyzed measure. However, it seems that the visual representation of these indicators could be also improved due to the fact that histograms can be also elaborated by other more simple tools and the tools used in this project should contribute with deeper analyses. These extra improvements are sought in the next chapter.
- Patterns (grouped). Finally, this view was used to compare also different groups of patients but now focusing on the amount of each type of activity. By means of this view, it is possible to group the same type of activities together and compare them with different groups of patients.

The rest of visualizations that were produced during the 2nd analysis phase but not used in the final meeting were:

- The "all activities" and "most frequent activities" views. These views contain more specific information that was required by the medical specialists but, during the final meeting, more general results were intended to be shown to the doctors.
- Process mining results. In the method, the process mining results were intended to get process diagrams that enhance the process insights of the clinical pathways. This goal was achieved in the project where the method was developed. However, as proved during the analysis phases of this validation process, these results did not contribute in getting new process information (i.e. the models shown in APPENDIX J do not clarify or add new knowledge from the one that can be gaining by analyzing the patterns sequential view in Figure 13). The benefit of using a process mining tool, as has been mentioned in Section 2.1, is that by means of this tool it is possible to look inside healthcare processes. Moreover, this tool offers more than one hundred and twenty plug-ins for analyzing and mining process information. This validation process has only tested three of them. Therefore, it seems that the capabilities of process mining, especially of the ProM tool, have not been completely exploited. Thus, some extra analyses are carried out with this tool in the following chapter.
- Handover of work. This view is not included in the method and it was also developed during this graduation project. This organizational aspect in a healthcare organization has proved to be important information for the medical specialists. In fact, during this validation process, the medical specialists at the AMC required this information to know the collaboration and handover of work between departments. This view was not shown during the final meeting because it also presents detailed information that requires a lot of time to analyze. Although this visualization received a positive feedback during the preliminary meeting, this view can also be improved with

further analysis because it only considers the relationships of departments that are produced under direct succession of work. This aspect is also investigated in the next chapter.

Now that the final comments of the views and tools used in the final stages of the method have been mentioned it is possible to present the final recommendation in using each specific view and tool for future healthcare process analysis. Summarizing, Table 1 presents all the views and analyses performed during the analysis phases of this validation process (preliminary analysis and the 2nd analysis phases). It shows in the second column if each view was used during the validation process at the AMC. Additionally, it gives, in the third column, the recommendation in using the view and an explanation in the following column of the table. The tools used for the analyses mentioned in the first column of the table were developed using mainly the MagnaView tool. Only the process mining results were calculated with the ProM tool. The *italic* analyses are not yet included in the method but developed during this validation process to fulfill process information requirements of the AMC. The criteria for recommending each analysis are based on two different aspects:

- First, if the analysis provided useful process information during the validation process.
- And second, if the analysis was not used during the validation process, then it was checked whether it was a useful analysis during the previous project that develops the method.

ANALYSES		Used in validation?	Usage recommendation?	Comments
	1. All activities	✓	✓	It is recommended because it gives a general overview of the most frequent activities in the processes
	2. First activities	✓	✓	It was only used in the preliminary analysis phase, however it provides clear information on how patients enter their process
	3. Last activities	✓	✓	It was only used in the preliminary analysis phase. It provides clear information on how patients leave their process. However, in order to be useful, this analysis must only be used for patients which processes are finished or completed
	4. Patterns (sequential)	✓	✓	The most important view during the validation project. It provides with process insights and the level of standardization in the medical processes
	5. Patterns (grouped)	✓	✓	Useful in providing insights about the number of activites per type. Additionally, it was used to compare activities among different groups of patients
	6. Process mining results	✓	?	Nevertheless the plug-ins suggested in the method provide important process information during the development of the method, in this validation process did not contribute to increase the process knowledge. However, the tool is powerful and extra analysis will be conducted in the following chapter to determine which plug-ins provide useful information.
	7. Centered activity with patterns (sequential)	✓	✓	Once a centered activity is identified this view can provide important process information. In this validation process it contributes to establish the level of the standardization of the processes
	8. Centered activity with patterns (grouped)	✗	✓	This view was not required by the medical specialists in the validation process. However, it can provide more process insights for healthcare analysis and it can be used also to compare activities for different centered activity views
	9. Centered activity with causal relations (1 step)	✗	✓	Not developed in this validation process. However, it could provide more process insights for healthcare analysis in knowing which activities were the preceding and subsequent activity of the centered activity. This analysis was useful during the development of the method
	10. Centered activity with causal relations (3 step)	✗	✓	Not developed in this validation process. However it can provide information about local patterns in the method by showing the 3 activities performed before and after certain centered activity
KPIs	*11. Averages times between certain activities*	✓	✓	It gives a a clear representation of the times by means of histograms. The distribution and most frequent values are easily identified
	12. Ocurrences of certain activities	✓	✓	It gives a a clear representation of the ocurrences of activities by means of histograms. The distribution and most frequent activities are easily identified
Extra analysis	*13. Handover of work*	✓	?	It provides useful information about the handover of work and collaboration between departments. However, it only takes into account the direct sucession of work. Further analysis could investigate it for improvements
	14. Patterns per day	✓	?	It was only used in the preliminary analysis phase. It is possible to identify patterns but not really easily and clearly. Further analysis could investigate it for improvements
	15. Activities per day per patient	✓	✓	This view provides process knowledge related with the activities performed by patients per date in the hospital. Additionally it serves in this project to calculate extra performance indicators

Table 1: Summary of the views and analyses developed during the entire validation process

The next section presents the evaluation of the medical specialists focusing on their opinion about the usefulness of the tool.

4.2 Evaluation of the tool from the different stakeholders in the hospital

During the final meeting phase of the validation process, the results were shown first to the leading manager of the process innovation group at the AMC. According to his feedback, the expectations of the project were exceeded. He understood the results, found them interesting, and he recommended a second presentation with the staff of this department so they could see what the method can do. He mentioned that this analysis views present important process insights that could be used to assess certain specific aspects of clinical pathways. As recommended by him, the same presentation was shown to a small group of the healthcare process innovation department at the AMC. This group could also recognize the benefits of the tools and method for healthcare process analysis.

The second final presentation was held with a small group of medical specialists and doctors. During this presentation, the results were not completely understood by the medical actors. The presentation contained exactly the same results presented to the previous group. However, these stakeholders were mainly looking for some parameters or baseline to compare their results. Thus, the discussion generated after the presentation was not process related. Moreover, this audience found the results of the performance indicators more interesting than the visualizations that show the level of standardization in the clinical processes. The differences in the feedback between the first and second presentation could be explained with what McCormack (2001) defined as Business Process Orientation (BPO). According to the author, the BPO makes an organization focuses more on processes. This focus emphasizes that the work should be reported to the customers inside the process. Thus, this way of thinking makes the actors care for the overall healthcare process and, in consequence, care for all the processes inside a certain department because the relationships of activities inside a hospital are fully interconnected. Additionally, the author also highlights that a strong BPO strengths the commitment with process improvement. This BPO is definitely present in the personnel of the process innovation group but it is not so common among medical specialists. Then, the method should also consider this aspect in order to present useful process information for medical stakeholders in healthcare organizations. For that reason, the improvements of the method proposed in Chapter 6 should also examine the possibility to present process related information that could enhance the process awareness among doctors.

Despite the perceived feeling during the presentation to the medical specialists that the results did not have the expected effect, at the end of the second presentation this group of actors also agreed in recognizing the benefits of the tool and method for healthcare process analysis. Therefore, it can be said that the method provides useful healthcare process information. However, as mentioned in Table 2, where some recommendations in all the steps of the method are given, some extra work has to be done in order to improve the final results of the method. The recommendations of Table 2 are taken into account for the formulation of the resulting method in Chapter 6. It is important to mention that the lay out and information in Table 2 are similar as the ones presented in Table 1 where first, it is said if each step was used during the validation project and then, if it is recommended to have in the method. Additionally, the criteria used for recommending each steps are based on the same two points explained for Table 1.

As explained in the previous paragraphs, the method has to improve the understandability of the results, especially by presenting results that can increase the BPO of the doctors. Thus, the second research goal of this graduation project has not been completely achieved. The analysis in Chapter 5 tries to entirely fulfill this research goal.

Finally, before concluding with the method it is important to check if the results of this validation process comply with the success factors established when developing the method. These success factors are fully explained in Section 2.3.1. Besides the characteristics already mentioned, the results should also be:

- Presented within limited time
- Process models should have a high fitness value
- The approach should be positively evaluated by the medical specialists and managers
- The analysis should focus on certain aspects of the treatment process

These success factors were all completed excluding the second one which states that the produced process models should have a high fitness value. The process models produced during this validation project (process models presented in APPENDIX F and APPENDIX J) had a fitness value < 40%. This value represents the number of cases that can be reproduced by the produced model. Hence, less than 40% of the cases can be reproduced in the produced models during this validation project. Therefore, Chapter 5 also investigates on which plug-ins can provide better fitness values for the data of the AMC.

STEPS OF THE METHOD		Used in validation?	Usage recommendation?	Comments
1. Build database	1.1 Contact data manager	✗	✓	Not done in this validation process because data comes from previous analysis project. However this activity must remain in the method
	1.2 Extract data (Define business goal, identify data warehouse, start data collection, and integration and consolidation of the data)	✓	✓	As mentioned in the previous chapter, the integration and consolidation of the data is necessary only if the data comes from different information data warehouses
	1.3 Transform data (Convert to MVN-format, pre-analysis and pre-processing of the data, and create database and import data)	✓	✓	This step is necessary to obtain the data in the right formats. The method should include the creation of the MXML file in this phase
	1.4 Load (Use ProM import tool and load data into MagnaView)	✓	✓	The ProM import tool usage should be in the "Transform data" step
2. Introduction session	2.1 Meet specialists	✓	✓	If possible, the analyst should be introduced to all the medical actors and managers of the analyzed department
	2.2 Determine users	✓	✓	This activity should be performed explicitly so the analysts know which stakeholder needs must be covered by the analysis
	2.3 Determine status of change	✓	✓	The knowledge of previous change projects in a treatment process could serve to know which activities to analyze or where to focus the analysis of a specific process
	2.4 Comunnicate / determine goals	✓	✓	Important activity to clarify the scope of the analysis
	2.5 Determine first set of KPIs and extra	✓	✓	Important activy to determine which activities should be analyzed
	2.6 Obtain information about HC program	✗	✓	This activity is essential but to get this knowledge more time is needed. It should be recommended to perform this activity after the preliminary meeting but before the preliminary analysis phase as an independent step
3. Preliminary analysis		✓	✓	Results shown in Table 1
4. Preliminary meeting	4.1 Introduction (Explain method, discuss initial results, and let actors interactively "play" with the tool)	✓	✓	Important to plan time for the actors to use the tool. This activity could help in producing more enthusiasm and better feedback on the analysis. Additionally, the method must suggest only one meeting including at least one member of each user role
	4.2 Selection (4.2.1. Select activities that do not belong in treatment program, select that are mandatory in treatment process, and determine outliers)	✓	✓	This activity should be performed by the analysts before the meeting because healthcare processes include a lot of activities and the time for this meeting is limited. Only the doubts of the analysts should be consulted in this meeting
	4.3 Closure (Determine follow-up steps (Extra KPIs and/or focus process analyses))	✓	✓	The follow up steps requiered by the stakeholders must be clear enough to the analysts
5. 2nd analysis		✓	✓	Results shown in Table 1
6. Final meeting	6.1 Present and discuss results	✓	✓	The results should be presented to the complete team in only one meeting
	6.2 Receive feedback on results	✓	✓	This feedback could be used to improve the method and determine its usefulnes
	6.3 Determine which goals were reached	✓	✓	This feedback could be used to improve the method and determine its usefulnes
	6.4 Determine follow up steps	✓	✓	This feedback could be used to improve the method and determine its usefulnes
7. Documentation	7.1 Project document	✓	✓	The structure of this document should depend on the questions that the analysis tries to answer
	7.2 Project file MagnaView	✓	✓	The MXML file should be also delivered in the hospital

Table 2: Summary of the steps developed during the entire validation process

4.3 Final recommendations and follow-up activities

After the final meeting at the AMC and with the gained experience during the complete project at the hospital some recommendations, besides the ones stated in Table 1 and Table 2, can be described in order to improve the results of this method for future analysis. The involvement with the medical specialists seems to be essential to have more enthusiasm and expectations over the project. In the case of this graduation project, the contact with the medical actors was restricted only for the meetings at the hospital. Due to different factors, it was impossible to work in this project at the location and this factor produced that some doctors did not know the goals and objectives of the project. Therefore, the attendance of medical specialists to the final meeting was only of three doctors. The feedback and discussion after the presentation could be considerably improved with the entire team of doctors at that moment. Additionally, this method was offered to the AMC for the validation process but they did not ask or look for it by themselves. In other words, they did not have a problem to solve by the application of the method in their hospital. Before this project they were not looking explicitly for this type of project in the gynecological oncology department, therefore, the actors were not so enthusiastic and involved with the project.

So far, the validation process has finished. Chapter 3 proved the reliability of the method by showing that it can be applied in a different time and healthcare environment. These time and environment were different than the ones where the method was developed but it also provided satisfactory results. Additionally, the relevance of the method was also shown with the received feedback explained in this chapter. Consequently, the first research goal of this project, related with the validation of the process, was achieved.

As established in the previous section, the second research goal was not completely realized yet. The part of determining the users of the analysis results is completed. These users, as proposed in Riemers (2009), must be the managers or process innovation personnel and the medical specialists because they are the stakeholders in a healthcare environment that can use these results to propose and make improvements in the medical processes. On the other hand, despite the overall feedback received during the last phase of the validation process was positive certain opportunities to improve the method were identified. Mainly, these opportunities are with the process mining tool for obtaining understandable process information. Moreover, this section has presented an analysis that justifies the extension of the research over the process mining tool in order to improve the method. Therefore, some extra analysis is intended to be carried out. The results are presented in Chapter 6. This extra analysis is proposed to improve the results of the method and to increase the usage of the ProM tool in order to utilize the benefits offered by this approach. As mentioned in Table 1 and Table 2, the views and tools proposed for further analysis in this chapter are: the activities per date per patient view, KPIs, the process mining results and the handover of work. Additionally, some extra analyses are carried out to investigate if more plug-ins could contribute to generate useful process information for healthcare organizations. Furthermore, the results should contribute to increase the process awareness among the medical stakeholders. The following chapter elaborates on these issues.

5. IMPROVEMENTS IN THE METHOD USING THE PROM TOOL

As has been explained so far, some extra analyses were justified in order to increase the understandability of the method. Additionally, the process mining results were not used during this validation process. This chapter elaborates on these extra analyses in the ProM tool in order to increase the usage of the process mining tool in the method for healthcare process analysis. According to Section 1.3, the validation approach for this extra analysis is to make the same analyses using both tools and then evaluate which one could represent the best option for the different stakeholders in a hospital. Up to now, mainly all the analyses have been performed using the visual analytics tool, MagnaView. Therefore, this chapter describes the efforts on repeating the analyses mentioned in Table 1 with the ProM tool. First, the most significant results obtained with the ProM tool are described. And finally, a criterion is established in order to choose the best option for each specific analysis.

5.1 Same analysis of the method using the ProM tool

This section describes the investigation done with the ProM tool in order to obtain the same results that were produced during the validation process with the MagnaView tool. This section shows the most significant results obtained with this analysis and it is divided according to each of the steps proposed, used and developed during the analysis phases in the validation process. It is important to mention that this section shows the analyses done with the ProM tool in order to get the same results as with the MagnaView tool. However, in some cases this goal was not achieved. Additionally, the research done for each of the following analyses was stopped when:

- The results are equal to the ones obtained with the MagnaView tool or
- No other plug-in was found that could produce the same results as the MagnaView tool.

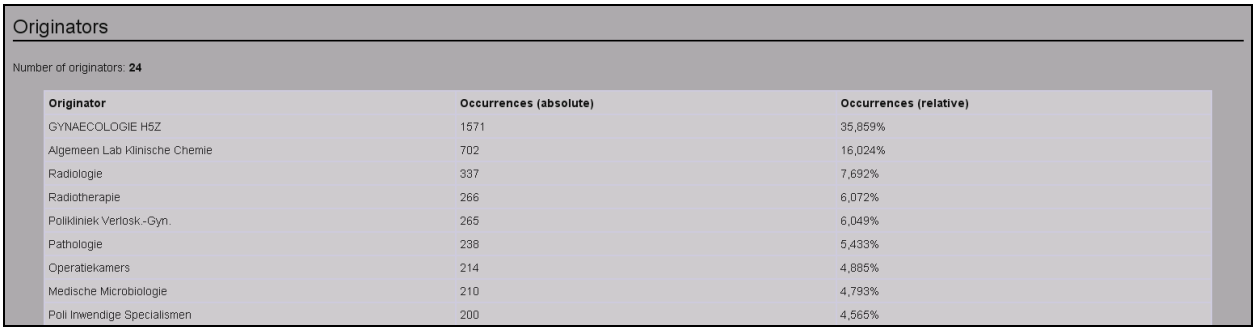
The log-file used for the coming analyses is the same log-file used to get the process mining results during the 2nd analysis phase of the validation process. This log-file contains the same pre-processing, aggregation and filtering activities described in Section 3.5 for the MagnaView file. As has been explained, the aim of these extra analyses is to produce the same results as the MagnaView tool so the tools can be compared, and this goal is only achievable if the analyzed data contains the same information for both tools.

5.1.1 All, first and last activities analyses

The results of the all, first and last activities analyses can be obtained using the *Log summary* plug-in. This plug-in is an analysis plug-in that provides statistical information about the log. This information includes the occurrences (absolute and relative) of the log events, the starting log events, the ending log events, and the originators. For example, Figure 21 shows how this information is presented for the most frequent originators in the DBC-code M13. The log events, starting log events, and ending log events information is produced by the *Log summary* plug-in in the same way as Figure 21.

The information presented in Figure 21 is the same as the one obtained with the MagnaView tool in Figure 14 for the DBC-code M13. However, no plug-in was found that could present the information in

the same way as Figure 14 where the comparison between three different DBC-codes is made. Thus, the information about the all, first, and last activities can be obtained with the *Log summary* plug-in in the ProM tool but this information has to be analyzed individually for each DBC-code because it is not possible to analyze and compare these three DBC-codes in one analysis.



Originators

Number of originators: **24**

Originator	Occurrences (absolute)	Occurrences (relative)
GYNAECOLOGIE H5Z	1571	35,859%
Algemeen Lab Klinische Chemie	702	16,024%
Radiologie	337	7,692%
Radiotherapie	266	6,072%
Polikliniek Verlosk.-Gyn.	265	6,049%
Pathologie	238	5,433%
Operatiekamers	214	4,885%
Medische Microbiologie	210	4,793%
Poli Inwendige Specialismen	200	4,565%

Figure 21: Originators analysis obtained using the Log summary plug-in of the ProM tool

As was explained during the 2nd analysis phase of the validation process, a view was developed in order to show the most frequent activities in the surgery and radiotherapy departments (the view for the most frequent radiotherapies is presented in Figure 15). This information can also be obtained from the *Log summary* plug-in. However, it is necessary to filter the needed information for such an analysis. One way to do it is to use the *Originator log filter* localized in the advanced filter options of the ProM tool in order to produce a log where only the activities of the surgery or radiotherapy department are included. Then, the *Log summary* plug-in can present the information of the particular activities inside a specific department, as shown in Figure 21. Additionally, it is possible to present this information in a bar-chart using the *Basic performance analysis* plug-in. Figure 22 presents the most frequent radiotherapy activities for the DBC-code M13. Still, again it is not possible to compare the results of M11, M13 and M16 in only one visualization, as Figure 15.

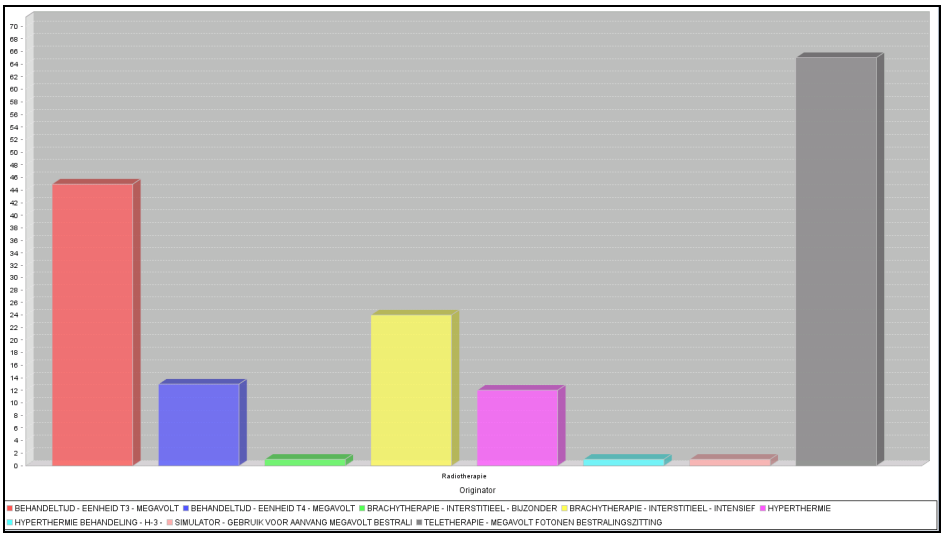


Figure 22: Most frequent radiotherapy activities for M13 produced with the ProM tool

5.1.2 Patterns (sequential)

The pattern (sequential) analysis was performed using the *Dotted chart analysis* (Song & Aalst 2007a) plug-in. According to Song & Aalst (2007a), the basic idea of the dotted chart analysis is to plot dots for each activity, according to the time when they were executed.

As has been explained in previous sections, the data for this analysis was extracted from the billing system of the hospital. This information system captures process-related information. However, the timestamp needed and used in the system is the day when an activity is performed. Therefore, at the moment of using this analysis plug-in a difficulty emerged due to the fact that for the activities that were performed for the same patient on the same date only one colored dot was visible because this dot hides the dots representing activities that occur on the same day too. Thus, in order to produce this analysis, some extra pre-processing activities were performed to the data. These extra activities were not done in the ProM tool. The data was exported into a spreadsheet application (Microsoft Office Excel). From there, consecutive minutes were added to the timestamps of the data that were performed on the same date for the same patients. Figure 23 shows the resulted visualization from the *Dotted chart analysis* plug-in. This figure contains the same information as Figure 13 where the first three months of activities for the patients that are since the beginning of their process are shown.

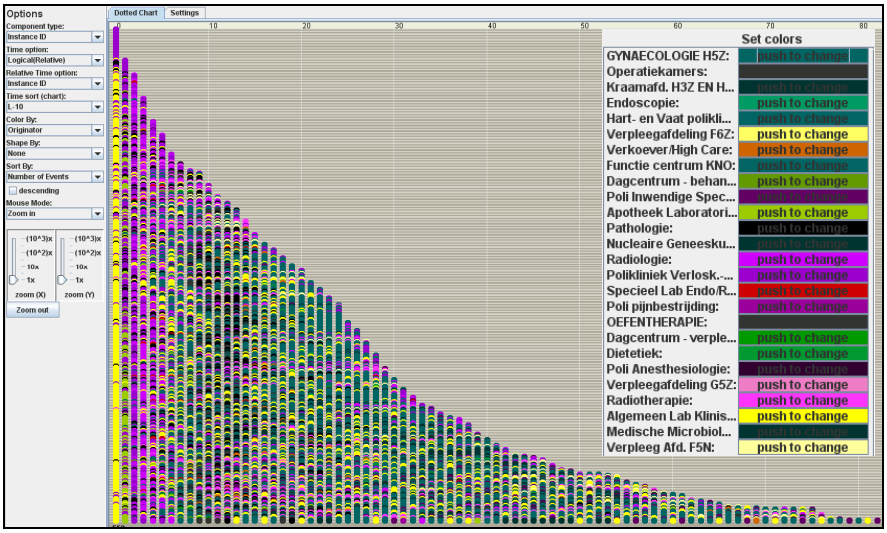


Figure 23: Pattern (sequential) analysis using the ProM tool

Furthermore, this plug-in also shows some performance indicators. These indicators present information of each specific component and general information of all the components included in the chart. The performance measures are the number of dots, the time of starting and ending of the events, and the average, minimum and maximum interval for the components. Moreover, using this plug-in it is possible to select the data based on their visual representation and use it for further analyses. These performance indicators are also possible to calculate and show using the MagnaView tool. However, these indicators are not presented automatically by the tool so they must be developed.

5.1.3 Patterns (grouped)

Similar as the patterns (sequential) analysis in the ProM tool, the pattern (grouped) analysis also needed some extra pre-processing activities in order to obtain the results with the ProM tool. Figure 24 shows the obtained visualization after performing the pre-processing activities.

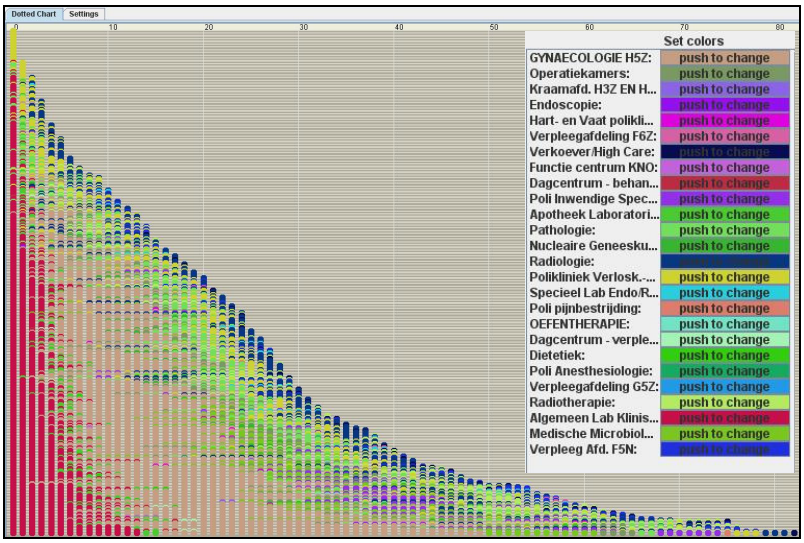


Figure 24: Pattern (grouped) analysis with ProM tool

The pre-processing activities were the modification of the timestamps in the events per patient. Detailedly, the same spreadsheet application used in the previous analysis was utilized in this section. With this tool, the data was sorted first per patient and then per type of activity. Then, the dates of the activities were added to the dataset in an ascend order so the same type of activities remain together in the visualization. Figure 24 presents the same information as shown in Figure 20 but with the ProM tool.

5.1.4 Process mining models

It is important to mention that a lesson learned during this research was that in order to produce useful process models from the mining plug-ins of the ProM tool it is necessary to reduce the dataset into the smallest possible groups of patients and analyze these groups separately. After the validation project at the AMC and during this extra analysis with the ProM tool it was noticed that the process models produced during the validation project contained the six different DBC-codes of patients. Therefore, the resulted process models contained information of six different treatment processes.

In this section, in order to produce the smallest possible datasets the log-file utilized during the 2nd analysis phase of the validation project was divided according to the DBC-code using the *LTL checker* plug-in. The *LTL checker* plug-in is based on a Linear Temporal Logic (LTL) language and it combines this language with a standard XML format to store event logs. Given an event log and a LTL property, the LTL checker verifies whether the observed behavior matches the (un)expected / (un)desirable behavior (Aalst et al. 2005). Thus, by using this plug-in six different log files were created, one per analyzed DBC-code (M11, M12, M13, M14, M15 and M16). Then, the *SOM* plug-in was used in order to cluster

different groups of patients inside the same DBC-code. For the details in the number of clusters and patients per cluster obtained in each DBC-code see APPENDIX L. Each of the clusters produced by the *SOM* plug-in was saved as a different log-file. Finally, different mining techniques were used over these new log-files in order to determine which mining plug-in could offer the best option for process diagrams in a healthcare environment.

An important aspect that should be considered in judging the usefulness of the models is their fitness measure. This fitness measure assesses the quality of the models by highlighting how many log traces can be replayed in the model (Alves de Medeiros 2006). There are different fitness measures implemented in the ProM tool and they measure different fitness aspects of the data. According to Alves de Medeiros (2006), a good fitness measure should be "complete and precise from a behavioral perspective". The completeness of a model indicates the number of traces that can be parsed in the model while the preciseness indicates how much extra behavior can be allowed in the model. The author recommends the usage of the *Extra Behavior Punishment* fitness measure because it measures both characteristics, preciseness and completeness. Hence, the mining models were compared, if possible, using this fitness measure.

The mining algorithms used and compared in this part of the analysis were the *Heuristic miner*, the *Genetic miner* and the *Fuzzy miner*. These algorithms were chosen because they produced heuristic net models which facilitates the understanding of a model (e.g. the *Alpha algorithm* produces Petri Net models and these models could difficult the understanding of the model among medical actors because besides the model the specialists needs to understand the meaning of the figures utilized in a Petri Net model). Next paragraphs describe the results on these three mining tools for one specific group of patients (57 patients) in the DBC-code M16, as produced by the *SOM* plug-in. This group of patients serves as a representative example, due to the fact that the results from the analyzed groups of patients were quite similar. Additionally, it is important to mention that the process models were extended with artificial start and end events to clearly identify the flow of activities in the model. These additions were done using the *Add Artificial Start Task Log Filter* and the *Add Artificial End Task Log Filter* in the advanced filter options of the ProM tool.

According to Zhou (2008), because of the iterative nature of healthcare processes, with several lab tests and repeated visits throughout the whole process, the *Heuristic miner* plug-in can have better results if the setting for the AND-threshold is changed to 10 and the setting for the length-one-loops is changed to 0.999. These changes were applied in this analysis. The resulted process model for our example model presented an *Extra Behavior Punishment* fitness measure of 0.8389. Then the *Genetic miner* was applied to the same log-file and the *Extra Behavior Punishment* fitness measure was 0.8767. Besides both models have fitness > 80%, there are some differences between both models and between the flows of activities through them. For the details in these models see APPENDIX M. Finally, the *Fuzzy miner* model was produced. This model is also presented in APPENDIX M. For this model it is not possible to obtain the *Extra Behavior Punishment* fitness measure in the ProM tool. Therefore, the comparison with the other models is difficult. However, this plug-in presents a different indicator called "log conformance". This measurement calculates the amount of events in the log that can be successfully replayed by the model. This measurement in more related with what Alves de Medeiros (2006) defined as completeness of the model. Due to the fact that no "preciseness" measurement can be calculated with the *Fuzzy miner* plug-in

we have to look only at the completeness measurement. For the case of the analyzed log-file the *Fuzzy miner* log conformance was 0.8244.

As has been shown, the three analyzed plug-ins produced satisfactory results in terms of the fitness of the model. Thus, in order to choose a plug-in for healthcare process analysis different aspects must be considered. Moreover, the three models can be presented to the different medical stakeholders so they can evaluate them and choose the best options or focus the analysis in some specific parts. Additionally, the models produced for the smallest possible group of patients are simpler than the ones that contain different treatment processes. The advantages in each of these three plug-ins are:

- According to Weijters et al. (2006), the *Heuristic Miner* plug-in only considers the order of the events within a case and not the order of events among cases. Thus, this plug-in is a practical applicable mining algorithm which can be used to express the main behavior registered in an event log and not all details and exceptions.
- According to Alves de Medeiros (2006), the *Genetic miner* plug-in can deal with different difficulties such as parallelism in the data, loops, non-free-choice and invisible tasks. Moreover, the author states that the *Heuristic Miner* plug-in mainly works based on binary relations which makes that the non-free-choice constructs cannot be captured. Thus, if non-free-choice activities are identified in the data, then it is probably better to use this plug-in.
- According to Gunther & Aalst (2007), when people are free to execute anything in any given order they will usually make use of such feature. The authors showed that the traditional desire to model the complete behavior of a process in a precise manner conflicts with the original goal of providing understandable and high-level information. Therefore, the *Fuzzy miner* offers multi-perspective options which make possible to simplify and aggregate data according to the necessities. Additionally, the *Fuzzy miner* produces models where the most frequent path is represented with ticker arrows, helping in identifying the most frequent path in the data. Another option found in the *Fuzzy miner* plug-in is the animation of the model. This animation offers a dynamic view of the process by replaying the log in the model.

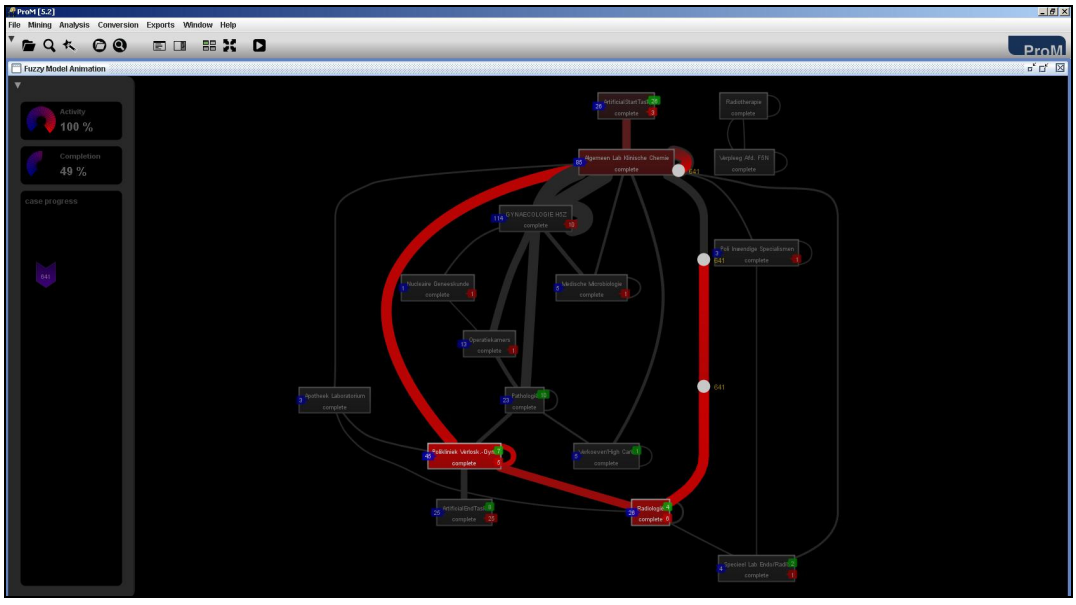


Figure 25: Fuzzy miner model animation

Additionally, the animation shows cases flowing through the model and the most frequent taken paths are highlighted (Mans et al. 2008a), as is shown in the screenshot of this animation in Figure 25. This animation could be shown to the medical specialists during the different meetings throughout an analysis project. The animation should probably contribute in a better understanding of process diagrams and to analyze the activities inside the clinical pathways with a process orientation. Thus, the process awareness can be improved among the medical actors and, in consequence, the comprehension of the overall results of this analysis method.

5.1.5 Patterns in the dataset

As has been shown, the *Performance sequence diagram analysis* plug-in can give information about the different patterns found in the dataset. Therefore, now this plug-in is used again but for the new groups of patients formed and explained in the previous section. Again, the group of 57 patients of the DBC-code M16 is explained in detail as a way to show results. The patterns were calculated for more groups of patients and the results depend on each specific case. However, this plug-in proved to be a good option in analyzing the different patterns among groups of patients.

The group of 57 patients of the DBC-code M16 resulted in 35 different patterns with 74,2% of unique patterns and the most common pattern is repeated 8 times. These results confirm the previous findings on the large amount of variation in the clinical pathways, even inside of this small group of patients.

As explained in the preliminary analysis phase (Section 3.3), one visualization analysis was developed in order to identify recurrent patterns in the data. The name proposed for such visualization was “Patterns per day”. This view was not easily interpreted by the medical specialists despite that they recognized its importance; therefore, some extra research was performed in order to find an option for this analysis using the ProM tool. Again, the *Dotted chart analysis* plug-in can also serve as an option to find patterns in the dataset. For example, the resulting visualization for this analysis in the group of 57 patients in M16 is shown in Figure 26. This figure also had the same overlapping of dots in the activities performed on the same date by the same patients. Again, the same pre-processing activities in the data of adding minutes to the timestamps were performed, as in Figure 23 and Figure 24. Additionally, as shown in the left side of Figure 26, the setting used for the time option in this figure was the logical option which sorted events based on their sequence, thus, the overlapping of events was solved.

Some recurrent patterns can be identified by analyzing this picture. In the MagnaView visualization (Figure 11) it was possible to select the patterns according to the number of activities executed per day. For example, if the patterns of 2 activities per day wanted to be analyzed, first the view was changed to only present the combination of two activities per day in the patients, and then the activities that wanted to be analyzed were filtered. In Figure 26, it is possible to have an overall visualization of the patterns. Additionally, with detailed analysis it should be possible to identify and visualize recurrent waiting times between patterns so the calculation of these waiting times can be executed in further analyses.

In conclusion, this combination of plug-ins could be a good option in finding patterns inside the dataset, the *Performance sequence diagram analysis* combined with the *Dotted chart analysis* visualization previously shown.

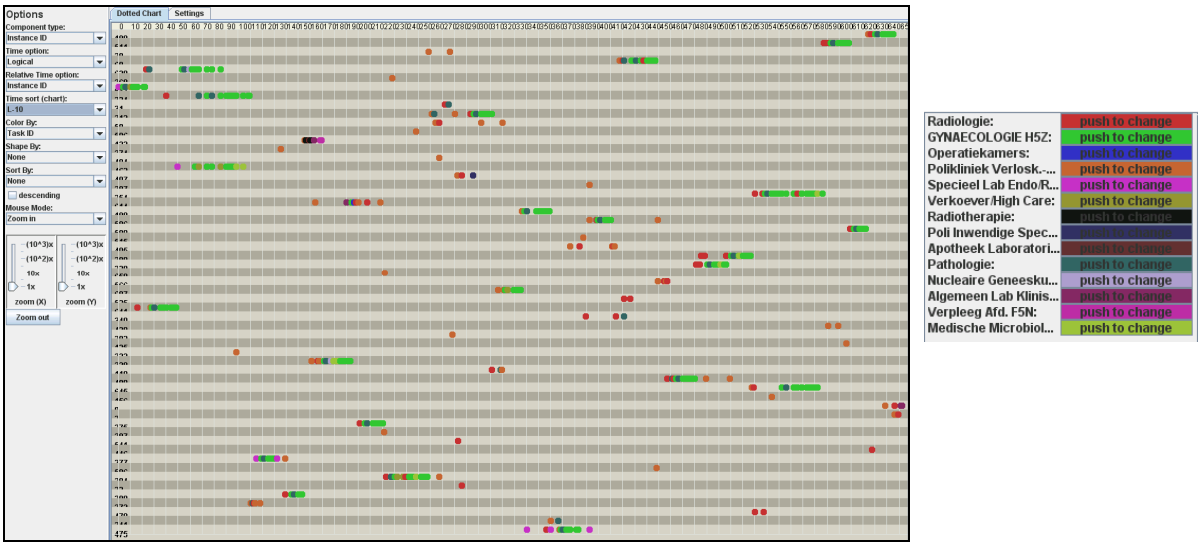


Figure 26: Patterns in the dataset using the ProM tool

5.1.6 Handover of work

The organizational related-information has proved to be an important aspect that should be included in any healthcare process analysis due to the fact that, during the validation process, the different stakeholders at the AMC ask for information on the handover of work and the collaboration between departments. The *Social network miner* plug-in can be used for such an analysis because it provides several social network analysis measures (Song & Aalst 2007b), like the handover of work.

The analysis measure of handover of work in the *Social network miner* plug-in provides a matrix with the information of the percentage of handover of work to each activity. Additionally, it is also possible to apply the *Analyze social network* plug-in to these results in order to obtain a view like the one shown in Figure 27. These two plug-ins offer different benefits due to the fact that the *Social network miner* presents the results and the *Analyze social network* its visualization. Figure 27 shows the social network for the activities in the DBC-code M16. The oval shape of the nodes in the graph expresses the relation between the in and out degree of the connections between those nodes. A higher proportion of ingoing arcs leads to more vertical oval shapes while higher proportions of outgoing arcs produces more horizontal oval shapes (Alves de Medeiros & Weijters 2006). Additionally, this plug-in provides an extra benefit in the analysis compare to the visualization developed using the MagnaView tool. In this ProM plug-in, it is also possible to not only consider direct succession of work but also indirect succession using the causality option. This option is part of the settings in the *Social network miner* plug-in. The complete list of options that can be modified in the settings of this plug-in are: to consider causality, to consider multiple transfers within one instance, and to consider only direct succession of work which gives the option to change the weights in the activities if the option is not selected.

Figure 27 is comparable with the data visualized in Figure 17 for the DBC-code M16. As noticed, Figure 27 presents more understandable information about the handover of work among departments in the treatment process M16. Additionally, the ProM tool produced similar results as the ones shown in Figure 27 for the DBC-codes M11 and M13.

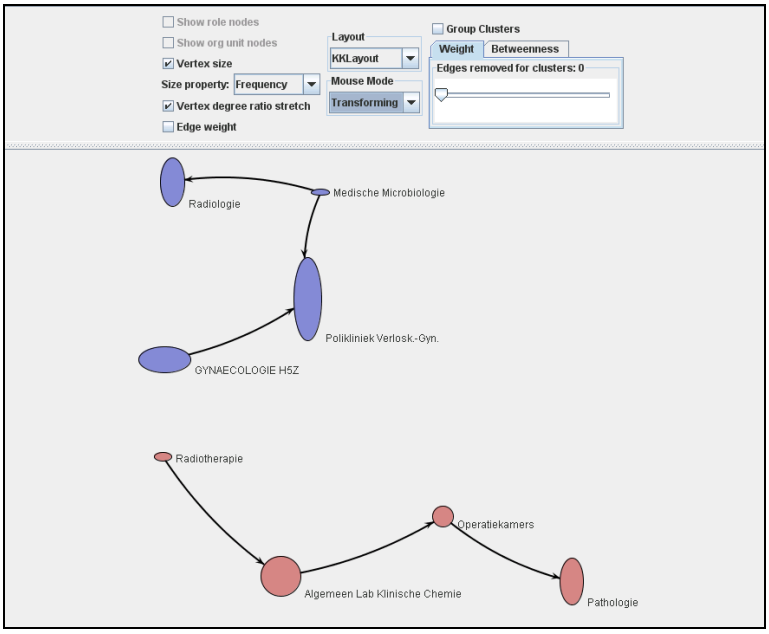
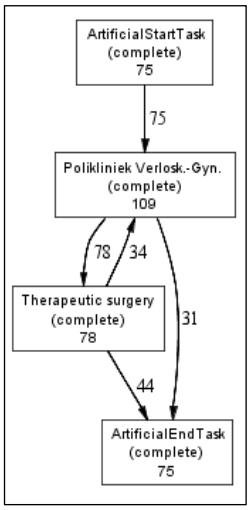


Figure 27: Handover of work visualization for patients in M16 using the ProM tool

5.1.7 Performance indicators

Some performance indicators can be automatically obtained with the ProM tool. For example, the *Performance analysis with Petri Nets* plug-in provides the average, minimum and maximum throughput time of cases, the average waiting time and sojourn time for each task, and the time between two tasks. Additionally, it also can provide information on bottlenecks and routing probabilities. Because during the validation process the AMC put more emphasis on the performance indicators related with the time between two different activities, this plug-in was investigated to obtain the KPIs indicated by the AMC.

Figure 28: Genetic model



For example, in order to obtain the average time between the patients first appointment with the doctors and their first therapeutic surgery activity in the DBC-code M13 (the log-file used for this analysis is the log-file that contains patients in the first three months of treatment that are since the beginning of their process in the DBC-code M13) the next process was followed: First, using the *LTL checker* plug-in, the patients in the DBC-code M13 were filtered and a different log was created. From this log, the therapeutic surgeries were identified and renamed as "Therapeutic surgery" with the *Remap Element Log filter* situated in the advanced filter options of the ProM tool. Then, the patients with "Polikliniek Verlosk.-Gyn." and "Therapeutic surgery" activities were the only ones kept in the log-file due to the fact that the time between these activities was

being looked for. These filtering activities were done using the *Event Log filter* in the advanced filter options of the ProM tool. The next activity was to eliminate duplicate activities so the calculation of this performance indicator can be done between the first appointment of patients and their first therapeutic surgery. These duplicate activities were eliminated by using the *Duplicate task filter* also situated in the advanced filter options of the ProM tool. Finally, the last step was to create a Petri Net model in order to utilize the *Performance analysis with Petri Nets* plug-in and calculate the average time between the activities. To do so, the *Genetic algorithm* plug-in was used because it was the mining tool that provides the best fitness measure (fitness of 0,975). The produced model is shown in Figure 28.

Figure 28 shows that 78 cases passed from a "Polikliniek Verlosk.-Gyn." activity to a "Therapeutic surgery" activity. However, only 75 cases entered into the model. As is shown in Figure 18, this value of 75 cases suits with the value obtained with the MagnaView tool for the M13 group of patients with a therapeutic surgery. Moreover, Figure 28 also shows that 34 patients returned to an appointment with their doctor but only 31 went then to the end task. By carefully analyzing the input data two important issues were checked in the log-file:

- First, that all the patients included in the analysis had a therapeutic surgery activity.
- And second, that all the patients that had an appointment followed by a surgery and then again an appointment went directly to the end task.

These rules were followed in all the patients except for three cases (patient-codes 284, 370 and 372) which after their first appointment, therapeutic surgery, and second appointment they returned again to a therapeutic surgery activity. No plug-in or filter option was found that could eliminate the second pair of these activities in these three patients. Thus, the average time between these two activities was expected to be different as the one obtained with the MagnaView tool. Just to corroborate the expectations, the *Performance analysis with Petri Nets* plug-in was used to calculate the time between these two activities. The results from this analysis showed that the average time for surgery in this group of patients is 30,72 days, with a standard deviation of 14,77 days. This data comes from analyzing the 78 cases showed in Figure 28.

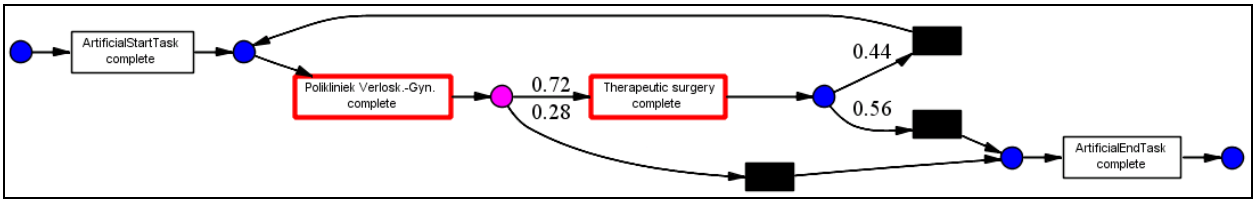


Figure 29: Performance analysis with Petri Net visualization

These values are indeed different compared to the values obtained with the MagnaView tool. In that analysis the results presented 75 patients, with an average of 26 days and a standard deviation of 16 days. Therefore, it can be concluded that this type of performance indicators cannot be correctly calculated with the ProM tool until an option is found that could eliminate the repetition of the two analyzed activities for the same patient. Nevertheless, the model presented in Figure 29 is a very simple model that can be easily interpreted by the medical specialists and this way of presenting this performance indicator can also contribute to increase the process awareness among the medical specialists, helping in achieving this goal in the analysis. However, it should be remembered that in order to use the ProM tool to calculate

performance indicators like this one, the analysts should be sure that no repetition of the two analyzed activities is present on the analyzed patients.

5.2 Criteria for using ProM or MagnaView

Section 5.1 presents different analyses made with the ProM tool. In some cases, the obtained results are the same as the ones obtained with the MagnaView tool. Thus, some criteria must be defined in order to delineate which approach is the best option for each specific case.

Vafaie et al. (2006) developed some selection criteria that answer the question of how to select the most appropriate intelligence analysis tool for analysis processes. According to the authors, the criteria are based on selection rules which must address issues raised by the customers. They developed a total of one hundred eight down-selection criteria, grouped according to seven dimensions. These dimensions are:

- User interface which relates to the quality of screens that enable users to interact with the system
- Architecture which describes the technical environment required for effective operation (hardware, platform, disk memory, etc.)
- Functionality. This dimension was divided in 8 categories: 1) Data analysis, 2) Extraction, transformation & load, 3) Search, 4) Categorization & summarization, 5) Link analysis, 6) Visualization, 7) Reporting / Collaboration, and 8) Deployment
- Technical support. Quality of customer support provided by the vendor
- Pricing. This dimension includes cost of software, training maintenance and technical support
- Installed customer base which determines the number and extent of existing major client implementations
- Company profile which describes the company's overall health

The ProM and MagnaView tools are going to be evaluated according to these criteria. This evaluation will be based only on the experience gained with both tools during this graduation project. Due to the limited time in this project no extra validation of these results was elaborated with more people who work with the tools. The main conclusions of these dimensions applied to the ProM and MagnaView tool are:

- The technical environment dimension describes the general architecture of a tool. This dimension is quite similar for both tools. Both tools operate over common platforms and no extra hardware is required for the tools.
- The technical support, pricing, installed customer base and company profile dimensions are different for the ProM and MagnaView tools. The main difference is that the ProM tool is a research tool while the MagnaView tool is a commercial tool. Thus, ProM offers articles and academic information on each plug-in of the tool and it is completely free. On the other hand, MagnaView offers support in the analyses and direct contact with their customers but it is not a free tool.
- The user interface and functionality dimensions are going to be evaluated in each of the individual analysis performed with both tools. This strategy is followed because, due to the plug-able nature of the ProM tool, each plug-in offers different characteristics and functions. Additionally, in this way it is possible to select the views and analyses on an individual level according to its possible usefulness in a healthcare environment. This evaluation is presented in Table 3.

Table 3 presents the user interface and functionality dimension conclusions over the analyses performed so far with the ProM and MagnaView tools. A detailed explanation of these conclusions can be found in APPENDIX N.

Views	produced in MagnaView?	produced in ProM?	MagnaView or ProM?	Comments
All, first, last and most frequent activities analysis	✓	✓	MagnaView	No visualization was produced with the ProM tool for these analyses. Thus, the functionality dimension is better with the MagnaView tool. Additionally, no visual comparison between different groups of patients is possible with the ProM tool.
Pattern (sequential and grouped)	✓	✓	MagnaView	Again, no comparison can be made with the ProM tool in the same visualization between different groups of patients. Additonally, the analyses in the ProM tool requiered more efforts in the pre-processing of the data.
Patterns in the dataset	✓	✓	ProM	The ProM tool offers clearer visualization of patterns. Moreover, the ProM tool can calculate the number of patterns in the dataset by only appliying one plug-in.
Process mining models	✗	✓	ProM	No process model can be calculated with the MagnaView tool.
Handover of work	✓	✓	ProM	The ProM tool offers results and visualizations that give a better understanding of the dependencies between departments by taking into account the causality in their relationships.
Performance indicators	✓	✓	MagnaView	For the KPIs calculated in this project, the MagnaView tool produced more accurate results. However, the ProM tool shows better process representation of the indicators and the automatic calculation of performance measures. If the difficulties explained in Section 5.1.7 are overcome, then this tool could be used to calculate performance indicators.

Table 3: Comparison between the views in the ProM and MagnaView tool

Based in the previous statements about the seven dimensions included in the criteria it can be concluded that if the differences of the tool in the technical support, pricing, installed customer base, and company profile dimensions are not an issue in selecting a tool, both tools should be used because they could offer some complementary results in a healthcare process analysis. If both tools can be obtained, then the suggestion would be to utilize the MagnaView tool to produce the all, first, last, and most frequent activities analysis plus the patterns (sequential and grouped) and the calculation on the performance indicators. On the other hand, the ProM tool provides better results in the patterns in the dataset analysis, the process mining models and the handover of work. Furthermore, if the limitation of pricing is not bridgeable, all these analyses can be made with the ProM tool but with the limitations explained so far in this chapter. It is important to mention that the centered activity view analysis and the activities per day per patient analysis are planned to be included in the method. However, these analyses can only be developed using the MagnaView tool because no plug-in in the ProM tool was found that can provide this information.

To finalize this chapter it is important to mention that not all the plug-ins offered by the ProM tool were investigated during this project to produce healthcare process information. The research was limited to the plug-ins described in this chapter. Only some plug-ins were investigated due to the fact that the goal was to obtain the same results as in the MagnaView tool in order to compare them and propose the best option. Thus, once the results were achieved, the research in the ProM tool stopped. With the analyses developed so far and the recommendations explained in Table 3 it can be said that the second research goal of this graduation project has been achieved due to the fact that the understandability of the results was taken into account for choosing the best tool on each analysis. Moreover, the third research goal of this graduation project was also completed with the analyses developed in this chapter. These analyses included the extension / improvement of the method as explained in the selection of tools of Table 3. Next chapter presents the resulting method for healthcare process analysis.

6. RESULTING METHOD

The main goal of this graduation project is to develop a method for healthcare process analysis. This goal was sought by a validation process of an existing method that uses process mining and visual analytics tools for healthcare process analysis. This validation project was extended with some extra analysis in order to produce useful process information to healthcare organizations. This chapter shows the resulting method produced after joining all the conclusions and recommendations so far mentioned. This new method is presented in the following way: First, the activities of the method as proposed by Riemers (2009) are shown; second, some statements are explained in order to justify specific changes / improvements to the method; third, the new method is presented by listing the consecutive order of all the recommended activities for each phase of the method; and finally, the most important issues on each phase are explained.

Table 4 presents the list of activities for the first 2 phases of the method as proposed by Riemers (2009).

1. Build database	1.1 Contact data manager	
	1.2 Extract data	1.2.1 Define business goal
		1.2.2 Identify the datawarehouse
		1.2.3 Start data collection
		1.2.4 Integration and consolidation of the data
	1.3 Transform data	1.3.1 Convert to MVN-format
		1.3.2 Pre-analysis and pre-processing of the data
		1.3.3 Create database and import data
	1.4 Load	1.4.1 Use ProM import tool
		1.4.2 Load data into MagnaView
2. Introduction session	2.1 Meet specialists	
	2.2 Determine users	
	2.3 Determine status of change	
	2.4 Communicate / determine goals	
	2.5 Determine first set of KPI's and extra	
	2.6 Obtain information about HC program	

Table 4: First 2 phases of the method as proposed in Riemers (2009)

As is shown in Table 2, all the steps proposed in the first two phases of the method are recommended to remain in it. However, Table 2 also mentions some recommendations based on the experience gained during this project in order to improve the results on each phase. Therefore, by following those recommendations, some changes are proposed. These changes are mainly changing the order of some activities and adding some new activities that can help in accomplishing the goals of these two phases. The proposed changes for these phases of the method are:

- As is shown in Table 4, the method starts with contacting the data manager. Then, the next activity is defining the business goal which is related with the activity of investigating how the data is going to be used. This second activity is more associated with knowing the usage of the data. Moreover, it is not an activity needed in the extraction part of this phase. The usage of the data can be established with the data manager. Therefore, these two activities could be grouped together in one new initial part for this phase.
- Then, the extract data part comes. If activity 1.2.1 of Table 4 (Define business goal) is grouped with the first activity of the method, then the remaining activities for this part are: identify data warehouse, data collection, and integration and consolidation of the data. As explained in the comments column of Point 1.2 in Table 2, if only one information system of the hospital is used

as the data warehouse (in the case of the validation process described in this graduation project it was the billing system), then the pre-processing and pre-analysis activities can begin just after collecting the data. On the contrary, if more than one system is used, then some integration and consolidation of data activities must be performed.

- The third part in the build database phase is the transformation of the data. This part starts with the creation of the MVN file. However, this activity should be the last one of this part. First, some pre-processing analysis must be done in order to create a database with the desired data and then the MVN and MXML files can be created.
- The last part of the build database phase must include the activities of loading the MVN and MXML files into their respective tool.
- When the build database phase has finished, the introduction session phase can start. However, as explained in the last part of Section 3.2, some work has to be done before this meeting in order to present the method to all the different stakeholders defined as the users of the results (managers or process-innovation personnel and medical specialists, as concluded in Section 4.3). Due to the limited time availability of these stakeholders in a healthcare environment, this meeting should be planned in advance in order to have, if possible, all these stakeholders involved since the beginning of the project. Therefore, some extra activities were added to this phase in order to prepare for the meeting. First, a presentation must be elaborated. And finally, one meeting should be scheduled if possible with all the different stakeholders.
- The activities proposed in Riemers (2009) during the introduction session remain the same. Except for the last one. This new method also proposes the creation of a new phase for obtaining the general domain knowledge of the analyzed department(s). As explained also in Section 3.2 and in the comments of Point 2.6 in Table 2, this activity cannot be performed during the introduction session because this session is intended to present the benefits of this project to the medical actors and not to increase the understanding of the processes in the analyst. However, this activity could save a lot of time and prevent mistakes during the analysis phases. Therefore, it must be conducted after the introduction session but before the analysis phases.

By following the changes stated in the previous list, Table 5 presents the proposed activities for the first 3 phases of the method. It is important to mention that the activity of integration and consolidation of data (activity 1.2.3 in Table 5) should be performed only if the data comes from more than one information system.

1. Build database	1.1 Data usage	1.1.1 Contact data manager
		1.1.2 Determine usage of data
	1.2 Extract data	1.2.1 Identify data warehouse
		1.2.2 Data collection
		1.2.3 Integration and consolidation of data
	1.3 Transform data	1.3.1 Pre-analysis and pre-processing of data
		1.3.2 Create database and import data
		1.3.3 Convert to MVN and MXML formats
	1.4 Load data into both tools	
2. Introduction session	2.1 Before meeting	2.1.1 Elaborate presentation of tools and method
		2.1.2 Schedule introduction session
	2.2 During the meeting	2.2.1 Meet specialists
		2.2.2 Determine users
		2.2.3 Determine status of change
		2.2.4 Communicate / determine goals
		2.2.5 Determine first set of KPIs and extra
3. Obtain domain knowledge		

Table 5: Resulting method – first 3 phases

Table 6 presents the list of activities for the analysis phases of the method as proposed in Riemers (2009).

3. Preliminary analysis / 5. 2nd analysis	3.1 All activities	
	3.2 First activities	
	3.3 Final activities	
	3.4 Patterns (sequential)	
	3.5 Patterns (grouped)	
	3.6 Process mining results	
	3.7 Centered activity with patterns (sequential)	
	3.8 Centered activity with patterns (grouped)	
	3.9 Centered activity with causal relations (1 step)	
	3.10 Centered activity with causal relations (3 steps)	
	3.11 KPIs	3.11.1 General information
		3.11.2 Overview
		3.11.3 Results
		3.11.4 Validation (process analysis)

Table 6: Analysis phases of the method as proposed in Riemers (2009)

Again, some recommendations were stated, mainly in Table 1, about how to change / improve the steps for the analysis phases of the method. The proposed changes are:

- As explained in Section 5.1.4, some pre-processing activities have to be performed to the data in order to produce smallest groups of patients. These new groups of patients are intended to produce more focus analysis with the tools. Therefore, some initial steps are proposed in this phase in order to have some files ready to be analyzed in the ProM and MagnaView tool for detailed analyses. These initial steps propose first the usage of the *LTL checker* plug-in in the ProM tool in order to create different log-files when more than one DBC-code is being analyzed. Each log-file should contain all the patients with the same DBC-code. Second, with the produced log-files, the *SOM* plug-in should be used to cluster different groups of patients. These new groups should be saved again as different log-files. Third, after identifying the patients for each of the formed groups with the previous analyses, these groups should also be formed in the MagnaView tool. The way to do this is by grouping patients using the *Map* option of this tool. And finally, if needed, the pre-processing activities for detailed analysis (e.g. filtering activities) must be carried out. All these activities can be grouped in an initial part for these analysis phases called “Preparation of data”.
- As explained in Section 5.2 (mainly in Table 3), the analysis tool that should be used for the first three steps proposed in the analysis phases of the method (In Table 6 – all activities, first activities and final activities analyses) should be produced with the MagnaView tool. Additionally, as mentioned in Table 1, the final activities analysis should only be developed for patients that have completed their treatment process. Furthermore, these three analysis steps could be group in an analysis part called “Analysis of activities” which could represent the initial understanding of the data.
- A third analysis part could be created in order to group all the pattern analyses so far made and recommended. This part of the analysis should contain four analyses: The pattern sequential (Figure 13) and pattern grouped (Figure 20) analyses which should be developed using the MagnaView tool, as stated in Table 3; the activities per day per patient (Figure 19) analysis should be added to the method due to the benefits offered by this view during this graduation project, as explained in Table 1. Moreover, this view has to be developed in the MagnaView tool due to the fact that no plug-in was found in the ProM tool that can produce it; and finally, the

pattern analyses using the *Performance sequence diagram analysis* and the *Dotted chart analysis* plug-ins of the ProM tool should also be added to the method. The usage and analysis produced by these plug-ins are fully explained in Section 5.1.5. Furthermore, these analyses should replace the patterns per day (Figure 11) analysis developed during the preliminary analysis phase of the validation process described in this project due to the fact that they provide clearer results. Therefore, the question mark placed in Point 14 Table 1 for this visualization has been answered.

- The fourth part for the analysis phases of the method should correspond to the centered activity views. As mentioned in Table 1, despite only one visualization for this analysis was used in this graduation project, all the visualizations as proposed in the method should remain in it. Moreover, as explained in Section 5.2, these analyses should be performed with the MagnaView tool due to the fact that no plug-in in the ProM tool was found that can provide this information.
- The process mining models, described in Section 5.1.4, could be developed using three different mining plug-ins of the ProM tool: the *Heuristic miner*, the *Genetic miner* and the *Fuzzy miner*. Section 5.1.4 also explains that these three plug-ins presented satisfactory results in producing models for small groups of patients. Additionally, the fuzzy animation in the *Fuzzy miner* plug-in could also be added to the method in order to increase the process awareness among the medical stakeholders. This animation is also explained in Section 5.1.4. Thus, by adding these analyses in the method the question mark put in Point 6 of Table 1 could also be answered in a positive way.
- As mentioned in Section 5.1.6, the organizational information proved to be an important requirement of healthcare process information during this project. Therefore, a handover of work analysis part should also be added to the method. Table 3 recommends the usage of the *Social network miner* and the *Analyze social network* plug-ins of the ProM tool for such an analysis instead of the handover of work visualization (Figure 17) developed with the MagnaView tool. Thus, the question mark added in Point 13 of Table 1 for this analysis was also answered by including these ProM analyses in the method.
- The KPI analysis should remain in the method. Some recommendations about the calculation of performance indicators with the tools are presented in Section 5.1.7 which concluded in Table 3 that for the KPI required by the AMC the MagnaView tool offers more accurate results.
- A final part in the preliminary analysis phase should also be added in order to: prepare a presentation with the most important results found during the analyses, prepare specific questions to the medical stakeholders about doubts to include/exclude data, and schedule the preliminary meeting so, as recommended in Section 3.4, this meeting can include a small group of people but be performed only once with the two different stakeholders defined as the users of the analysis results (managers or process-innovation personnel and medical specialists). Moreover, it is important to mention that the questions about which activities to include / exclude from the analysis were included in the method proposed by Riemers (2009). The author recommended executing these activities during the preliminary meeting phase. However, these activities could take a lot of time due to the large number of activities performed in a DBC-code and the purpose of this meeting is only to present results and clarify doubts in the analysis. Hence, this research was transferred to this analysis phase and during the preliminary meeting phase these doubts should only be clarified.

Table 7 shows the resulted method for the preliminary analysis and 2nd analysis phases included in the method. These activities, as the rest of the proposed activities in the different resulting method tables,

should be performed consecutively as presented in the tables. It is important to mention that the activities proposed in part 4.8 of the method should only be performed during the preliminary analysis phase of the method. Thus, the KPIs activity should be the last activity in the 2nd analysis phase of the method.

4. Preliminary analysis / 6. 2nd analysis	4.1 Preparation of the data	4.1.1 LTL-checker / separate DBC-codes
		4.1.2 SOM plug-in to each DBC-code
		4.1.4 Create maps in MagnaView for the different group of patients
		4.1.4 Pre-processing activities for detailed analysis
	4.2 Analysis of activities	4.2.1 All activities
		4.2.2 First activities
		4.2.3 Last activities
	4.3 Patterns	4.3.1 Patterns (sequential)
		4.3.2 Patterns (grouped)
		4.3.3 Activities per day per patient
		4.3.4 Patterns in dataset
	4.4 Centered activity	4.4.1 Sequential
		4.4.2 Grouped
		4.4.3 Causal relations (1 step)
		4.4.4 Causal relations (3 steps)
	4.5 Process mining models	4.5.1 Process models (Heuristic miner / Genetic miner / Fuzzy miner)
		4.5.2 Fuzzy animation
	4.6 Handover of work	4.6.1 Social network miner plug-in
		4.6.2 Analyze social network plug-in
	4.7 KPIs	
	4.8 Results	4.8.1 Elaborate presentation with the most important findings
		4.8.2 Prepare questions to stakeholders about doubts in the data
		4.8.3 Schedule the preliminary meeting

Table 7: Resulting method - analysis phases

Finally, Table 8 describes the activities for the last 3 phases of the method as proposed by Riemers (2009).

4. Preliminary meeting	4.1 Introduction	4.1.1 Explain (calculation) method	
		4.1.2 Discuss initial results	
		4.1.3 Let actor(s) interactively "play" with tool	
	4.2 Selection	4.2.1 Select activities that do not belong in treatment program	
		4.2.2 Select activities that are mandatory in treatment program	
		4.2.3 Determine outliers	
	4.3 Closure	4.3.1 Determine follow-up steps	4.3.1.1 Extra KPIs
			4.3.1.2 Focus process visualization / analysis
6. Final meeting	6.1 Present & discuss results		
	6.2 Receive feedback on results		
	6.3 Determine which goals were reached		
	6.4 Determine follow-up steps		
7. Documentation	7.1 Project document	7.1.1 Summary (max 1-2 pages)	
		7.1.2 General information	7.1.2.1 Goals
			7.1.2.2 Target group
			7.1.2.3 Data
		7.1.3 List of all activities	
		7.1.4 List of first activities	
		7.1.5 List of final activities	
		7.1.6 Most important patterns	
		7.1.7 Most important process mining results	
		7.1.8 Mot important centered view	
		7.1.9 KPIs	7.1.9.1 General information (calculation etc)
			7.1.9.2 Overview
			7.1.9.3 Results
			7.1.9.4 Validation
		7.1.10 Conclusion	
	7.2 Project file MagnaView		

Table 8: Last 3 phases of the method as proposed in Riemers (2009)

The changes proposed in these 3 last phases of the method are:

- As presented in Table 2, all the activities performed for these 3 final phases of the method should remain in it. However, in the case of the preliminary meeting phase some changes are going to be

proposed to adjust it to the changes done in the previous phase. Therefore, due to the fact that during the preliminary analysis phase a preparation of questions about which activities to include/exclude from the analysis, activities 4.2.1 and 4.2.2 of Table 8 can be joined into one activity that seeks to clarify the doubts encountered during the preliminary analysis phase.

- After the preliminary analysis phase, the 2nd analysis phase must be performed. As presented in Table 7, this phase includes the same analysis done during the preliminary analysis phase except for Part 4.8.
- In the final meeting phase some activities can be added to the method in order to prepare this meeting in the same way as the introduction session or the preliminary meeting. Therefore, again an initial part can be added to prepare the final presentation and to schedule the meeting. Then, during the meeting the four activities proposed in the method developed by Riemers (2009) remain the same.
- Finally, the last phase of the method proposes the content that should be included in the project document. However, in this graduation project was learned that the information presented in this document should depend entirely in the information requirements of the healthcare organization. Therefore, the content proposed in this document should be eliminated. Additionally, the method only considers delivering the MagnaView file. However, the MXML files should also be delivered to the healthcare organization for their archives.

5. Preliminary meeting	5.1 Introduction	5.1.1 Explain the results
		5.1.2 Discuss initial results
		5.1.3 Let actors interactively "play" with tools
	5.2 Selection	5.2.1 Clarify doubts about activities to include / exclude
		5.2.2 Determine outliers
	5.3 Determine follow up steps	
7. Final meeting	7.1 Before meeting	7.1.1 Elaborate presentation with the most important
		7.1.2 Schedule the final meeting
	7.2 During meeting	7.2.1 Present and discuss the results
		7.2.2 Receive feedback on results
		7.2.3 Determine which goals were reached
		7.2.4 Determine follow up steps
8. Final documentation	8.1 Project document	
	8.2 Project files (MVN and MXML files)	

Table 9: Resulting method - last 3 phases

The complete list of activities for each phase in the resulting method can be found in APPENDIX O. Next chapter elaborates on the final conclusions of this graduation project, highlighting also its main limitations and recommendations for future research.

7. CONCLUSIONS

The goal of this graduation project was to develop a method for healthcare process analysis that present useful and understandable process information in hospital settings. To do so, a validation process of an existing method (Riemers 2009) that uses process mining and visual analytics was performed for the gynecological oncology department of the AMC. This method was originally developed in a different organization and for different treatment processes. Furthermore, some extra analyses were performed in this graduation project in order to improve the understandability of the method. More specifically, these extra analyses were completed with the process mining tool due to the fact that, during the validation process, it was noticed that the benefits offered by this tool have not been completely exploited. Thus, not only the conclusions of the validation process were used to propose improvements in the method but also the results of the extra analyses, where the process mining and visual analytics tools were compared by producing the same analysis with the tools and, at the end, using a defined criteria to choose the best option for a healthcare process analysis. The resulting method shows more usage of the process mining tool. Therefore, the main goal of this graduation project was achieved by proposing a method for healthcare process analysis that could enhance the process-related information needed by these organizations in order to produce process improvements.

Some specific conclusions about the developed views and analyses were gained throughout this graduation project. In this section I will emphasize the ones that I found more important for future healthcare process analyses. At the beginning of this report it was said that healthcare processes are mainly composed by high volumes of activities and that usually these processes have a large amount of variation. This statement was fully observed during this graduation project. Just to make a clearer illustration about the level of variation found, it should be remembered that even with the smallest analyzed groups of patients, with patients that share the same characteristics in their treatment like the DBC-code and time in their treatment, around 75% of patients were found with a unique pattern (Section 5.1.5). In this environment, the “Patterns (sequential)” view has proved to be an essential part in any efforts for healthcare process analysis. From the results visualized in this analysis it was possible to state that certain level of standardization was present in the dataset even with the large amount of variation found in the data. Then, as proved during this project, the “Centered activity” analysis confirmed those findings. This way of presenting information undoubtedly clarifies the processes in the analyzed department.

Furthermore, the visualizations produced in the MagnaView tool should be implemented with options to change the timeframe or to filter activities in order to make them more interactive. These extra features can improve the understandability among medical stakeholders and enhance the probabilities to find more useful process information from the analysis. On the other hand, as far as it concerns with the process mining tool, the ProM tool has increased its usage with the resulting method described in Chapter 6. However, the plug-ins described in Chapter 5 are the only ones that have been investigated. Additionally, some analyses were not reproduced with the ProM tool due to the fact that no plug-in was found that can produce certain results (e.g. Centered activity analyses). Therefore, future research should focus in assessing the usefulness of the rest of plug-ins offered by this tool due to the fact that it has proved to be a useful tool in healthcare process analysis projects.

The role of the different stakeholders involved in this method has not been explicitly mentioned so far. Therefore, it is important to clarify this issue. First the users of the analysis results must be, as previously mentioned, managers or process-innovation personnel and medical specialists. These stakeholders have been selected due to the fact that they are the stakeholders identified as capable actors of producing healthcare process improvements in a healthcare environment. These stakeholders must be involved during the project mainly in the meetings proposed in the method. In these meetings they should provide feedback about the results in order to have a useful analysis project. Additionally, they could be also involved during the analysis phases in order to produce information that is really required by the healthcare organization for process improvements. The other actor needed for this kind of projects is the analyst. This stakeholder is the responsible of producing the results by using the process mining and visual analytics tools. The profile of this stakeholder does not have to include knowledge in the analyzed healthcare processes but must have the technical knowledge to use the tools and perform the process analysis proposed in this graduation project. These three different stakeholders are needed for this type of analysis.

A main limitation of the produced method is that the analyses proposed to be carried out in the ProM tool were not validated or shown to the different stakeholders at the AMC. Therefore, the adding of these analyses into the method was justified by my experience gained during this graduation project and by a positive feedback received from the developer of the original method who assessed the changes made in the method. Nevertheless, the results may need some adjustments from the medical stakeholders. Furthermore, due to time limitations of this graduation project some of the proposed analyses in the method that use the process mining tool were not again developed in the MagnaView tool to see if it can provide better results. For example, the ‘Patterns in the dataset” analysis was proposed to be performed using two different plug-ins in the ProM tool. Future research could try to find a visualization produced by the MagnaView tool that can show the results as those plug-ins emphasizing the most common patterns in the dataset. Moreover, the animation proposed to be included in the method must also be tested in a healthcare environment. This type of animation can improve the understanding of processes among the actors and stakeholders involved in a healthcare analysis project.

8. REFERENCES

- Aalst, W.M.P. van der, Hofstede, A., Weske, M. (2003). Business Process Management: A survey. In W. Aalst, A. Hofstede, & M. Weske (Eds.), *International Conference on Business Process Management (BPM 2003).* Vol. 2678, pp. 1-12. Springer-Verlag, Berlin

- Aalst, W.M.P. van der, Beer, H.T. de, Dongen, B.F. van (2005). Process mining and verification of properties: an approach based on temporal logic. BETA Working Paper Series, WP 136, Eindhoven University of Technology, Eindhoven, 2005

- Aalst, W.M.P. van der, Netjes, M., Reijers, H.A. (2007a). Supporting the Full BPM Life-Cycle Using Process Mining and Intelligent Redesign. *In K. Siau (Ed.)*: Contemporary Issues in Database Design and Information Systems Development, Chapter 4, pp. 100-132. IGI Global, Hershey, USA

- Aalst, W.M.P. van der, Reijers, H.A., Weijers, A.J.M.M., Van Dongen, B.F., Alves de Medeiros, A.K., Song, M., Verbeek, H.M.W. (2007b). Business process mining: An industrial application. *Information systems*. Vol. 32, pp. 713-732

- Aalst, W.M.P. van der, Van Dongen, B.F., Gunther, C.W., Mans, R.S., Alves de Medeiros, A.K., Rozinat, A., Rubin, V., Song, M., Verbeek, H.M.W., Weijters, A.J.M.M. (2007c). ProM 4.0: Comprehensive Support for Real Process Analysis. In J. Kleijn and A. Yakovlev, editors, Application and Theory of Petri Nets and Other Models of Concurrency (ICATPN 2007), volume 4546 of Lecture Notes in Computer Science, pages 484{494. Springer-Verlag, Berlin, 2007.

- Aalst, W.M.P. van der, Rubin, V., Verbeek, H.M.W., van Dongen, B.F., Kindler, E., Gunther, C.W. (2008). Process mining: a two-step approach to balance between underfitting and overfitting.

- Alves de Medeiros, A.K. (2006). Genetic process mining. PhD Thesis. Technische Universiteit Eindhoven, Eindhoven, The Netherlands, 2006

- Alves de Medeiros, A.K. & Weijters, A.J.M.M. (2006). ProM framework tutorial. Technische Universiteit Eindhoven. Eindhoven, The Netherlands. November 2006.

- AMC website – Research – The AMC. Retrieved on July 4th 2009. From http://www.amc.nl/?pid=2581

- Anderson, J.G., & Balas, E.A. (2008). "Information technology in primary care practice in the United States". In: Tan, J. Healthcare information systems and informatics: research and practices. Hersey, New York; 2008

- Anyanwu, K., Sheth, A., Cardoso, J., Miller, J., Kochut, K. (2003). Healthcare enterprise process development and integration. *Journal of research and practice in information technology*. Vol. 35, no. 2, pp. 83-98

- Balls, M., Blaauboer, B.J., Fentem, J.H., Bruner, L., Combes, R.D., Ekwall, B., Fielder, R.J., Guillouzo, A., Lewis, R.W., Lovell, D.P., Reinhardt, C.A., Repetto, G., Sladowski, D., Spielmann, H., Zucco, F. (1995). Practical aspects of the validation of toxicity test procedures. ATLA 23, pp. 129-147

- DBC website. Retrieved on August 15th (2009). From http://www.dbconderhoud.nl

- Elkhuizen, S.G., Burger, M.P.M., Jonkers, R.E., Limburg, M., Klazinga, N., Bakker, P.J.M. (2007). Using business process redesign to reduce wait times at a University Hospital in The Netherlands. *The joint commission journal on quality and patient safety*. Vol. 33, no. 6, pp. 332-341

- Gunther C.W. and Aalst W.M.P. van der (2006). A generic import framework for process event logs. In: J. Eder and S. Dustdar, Editors, Business Process Management Workshops Workshop on Business Process Intelligence (BPI 2006), Lecture Notes in Computer Science vol. 4103, Springer, Berlin, pp. 81-92

- Gunther C.W. and Aalst W.M.P. van der (2007). Fuzzy mining: adaptive process simplification based on multi-perspective metrics. In G. Alonso, P. Dadam, and M. Rosemann, editors, International Conference on Business Process Management (BPM 2007), volume 4714 of Lecture Notes in Computer Science, pages 328-343. Springer-Verlag, Berlin, 2007

- Hornix, P.T.G., (2007). Performance analysis of business process through process mining. Eindhoven University of Technology 2007.

- Keim, D.A., Mansmann, F., Schneidewind, J., Ziegler, H. (2006). Challenges in visual data analysis. *Information visualization. IEEE Comput. Soc. 2006*, pp. 6

- Lin, FR., Chou, SC., Pan, SM., Chen, YM. (2001). Mining time dependency patterns in clinical pathways. *International journal of medical informatics*. Vol. 62, pp. 11-25

- MagnaView website – Home. Retrieved on July 4th 2009. From http://www.magnaview.com/

- Mans, R.S., Schonenberg, M.H., Song, M., Aalst, W.M.P. van der, Bakker, P.J.M. (2008a). Application of process mining in healthcare – A case study in a Dutch Hospital. *A. Fred, J. Filipe, and H. Gamboa (Eds.):* BIOSTEC 2008, CCIS 25, pp. 425-438

- Mans, R.S., Schonenberg, M.H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., Aalst, W. (2008b) Process mining techniques: an application to stroke care. *S.K. Andersen et al. (Eds)*: IOS Press 2008. Pp, 573-578

- McCormack, K. (2001). Business process orientation: do you have it? Quality progress. Vol. 34 (1) ABI/INFORM Global. Pp. 51-58.

- Riemers, P. (2009). Process improvement in Healthcare: A data-based method using a combination of process mining and visual analytics. Master of Science graduation project. Eindhoven University of Technology

- Simoff, S.J., Bohlen, M.H., Mazeika, A. (2008). Visual data mining: an introduction and overview. *S.J. Simoff et al. (Eds)*: Visual data mining, LNCS 4404, pp. 1-12

- Song, M. & Aalst, W.M.P. van der (2007a). Supporting process mining by showing events at a glance. In K. Chari, A. Kumar editors, Seventeenth Annual Workshop on Inform*at*ion Technologies and Systems (WITS'07), pp.139–145, Montreal, Canada, December 8-9, 2007

- Song, M. & Aalst, W.M.P. van der (2007b). Towards comprehensive support for organizational mining. BETA Working Paper Series, WP 211, Eindhoven University of Technology, Eindhoven, 2007

- Song, M., Gunther, C.W., Aalst, W.M.P. van der (2008). Trace clustering in process mining. 4th Workshop on Business Process Intelligence (BPI 2008).

- Thomas, J.J., & Cook K.A. (2006). A visual analytics agenda. IEEE Computer graphics and applications. Vol. 26 (1), pp. 10 – 13, January / February 2006

- Vafaie, H., Brown, N.F., Truong, L. (2006). Methodology for the selection of intelligence analysis tools. Proceedings of the 18th IEEE international conference on tools with artificial intelligence (ICTAI'06). Pp. 8-15

- Weijters, A.J.M.M., Aalst, W.M.P. van der, Alves de Medeiros, A.K. (2006). Process mining with the heuristic miner algorithm. BETA working paper series, WP 166, Eindhoven University of Technology.

- Yang, WS. & Hwang, SY. (2006). A process-mining framework for the detection of healthcare fraud and abuse. *Expert systems with applications*. Vol. 31, pp. 56-68

- Zhou, W. (2008). Process mining: acquiring objective process information for healthcare process management with the CRISP-DM framework. Master of Science graduation project. Eindhoven University of Technology

9. APPENDIX A

This section presents the complete method for healthcare process analysis using process mining and visual analytics tools as proposed in Riemers 2009. The method is divided in seven main stages which are: build database, introduction session, preliminary analysis, preliminary meeting, 2nd analysis, final meeting and documentation. Each phase presents detailed and specific steps that must be done in order to perform a successful healthcare process analysis project. The method uses ProM and MagnaView tools to perform the analysis proposed in the analysis phases. The author of the method worked on a business case, and based on the results of this business case and the experience gained, the method, with all its steps, was designed. Therefore, the following are the recommended actions for a healthcare process analysis that uses process mining and visual analytics tools with the aim of getting useful information from the healthcare processes as proposed in Riemers (2009):

1. Build database
    - 1.1. Contact data manager
    - 1.2. Extract data
        - 1.2.1. Define business goal
        - 1.2.2. Identify the data warehouse
        - 1.2.3. Start data collection
        - 1.2.4. Integration and consolidation of the data
    - 1.3. Transform data
        - 1.3.1. Convert to MVN-format
        - 1.3.2. Pre-analysis and pre-processing of the data
        - 1.3.3. Create database and import data
    - 1.4. Load
        - 1.4.1. Use ProM import tool
        - 1.4.2. Load data into MagnaView

2. Introduction session
    - 2.1. Meet specialists
    - 2.2. Determine users
    - 2.3. Determine status of change
    - 2.4. Communicate / determine goals
    - 2.5. Determine first set of KPIs and extra
    - 2.6. Obtain information about HC program (domain knowledge)

3. Preliminary analysis
    - 3.1. All activities
    - 3.2. First activities
    - 3.3. Final activities
    - 3.4. Patterns (sequential)
    - 3.5. Patterns (grouped)
    - 3.6. Process mining results
    - 3.7. Centered activity with patterns (sequential)
    - 3.8. Centered activity with patterns (grouped)
    - 3.9. Centered activity with causal relations (1 step)
    - 3.10. Centered activity with causal relations (3 steps)
    - 3.11. KPIs

3.11.1. General information
3.11.2. Overview
3.11.3. Results
3.11.4. Validation (process analysis)

4. Preliminary meeting
   4.1. Introduction
      4.1.1. Explain (calculation) method
      4.1.2. Discuss initial results
      4.1.3. Let actor(s) interactively "play" with tool
   4.2. Selection
      4.2.1. Select activities that do not belong in treatment program
      4.2.2. Select activities that are mandatory in treatment program
      4.2.3. Determine outliers (gem +40% activities, bottom/top 20% patterns, patients who miss activities, patients with unacceptable activities)
   4.3. Closure
      4.3.1. Determine follow-up steps
         4.3.1.1. Extra KPIs
         4.3.1.2. Focus process visualization / analysis

5. 2nd analysis (with adjustments)
   5.1. All activities
   5.2. First activities
   5.3. Final activities
   5.4. Patterns (sequential)
   5.5. Patterns (grouped)
   5.6. Process mining results
   5.7. Centered activity with patterns (sequential)
   5.8. Centered activity with patterns (grouped)
   5.9. Centered activity with causal relations (1 step)
   5.10. Centered activity with causal relation (3 steps)
   5.11. KPIs
      5.11.1. General information
      5.11.2. Overview
      5.11.3. Results
      5.11.4. Validation (process analysis)

6. Final meeting
   6.1. Present & discuss results
   6.2. Receive feedback on results
   6.3. Determine which goals were reached
   6.4. Determine follow-up steps

7. Documentation
   7.1. Project document
      7.1.1. Summary (max 1-2 pages)
      7.1.2. General information
         7.1.2.1. Goals
         7.1.2.2. Target group
         7.1.2.3. Data

7.1.3. List of all activities
7.1.4. List of first activities
7.1.5. List of final activities
7.1.6. Most important patterns
7.1.7. Most important process mining results
7.1.8. Most important center view
7.1.9. KPIs
7.1.9.1. General information (calculation etc)
7.1.9.2. Overview
7.1.9.3. Results
7.1.9.4. Validation
7.1.10. Conclusion
7.2. Project file MagnaView

10. APPENDIX B

The recognized and filtered administrative and no process-related activities were:

Activities filtered out
190101 Bovenreg. Toesl. A101
190204 klasse 3A A204
190205 klasse 3B A205
Administratief tarief – eerste polikliniekbezoek
Klinisch tarief
Klinische kaart – anesthesie
Klinische kaart – inwendige geneeskunde
Klinische kaart – verloskunde en gynaecologie
Ordertarief
Patient niet verschenen ngiv -no show
Patient niet verschenen radiologie –no show
Patient kontact audiologie korter dan 1.5 uur
Verlosk.-gynaec. Jaarkaart kosten out
Verlosk.-gynaec. Aanv.Kart kosten out
Verlok.-gynaec. Korte kaart kosten out

Table 10: Administrative activities filtered out during the pre-processing activities

The complete list of activities taken into account for the analyses is presented in Table 11:

190021 KLINISCHE OPNAME A002	ECHO BEEN	OVARIUM - REDEBULKING OVARIUM CARCINOOM
190034 AFWEZIGHEIDSDAG A006	ECHO BEEN - LINKS	PACLITAXEL
190035 DAGVERPLEGING A007	ECHO BEEN - RECHTS	PARTUS - KLIN.OF POLIKLIN.ZND VOORBEH.OP MED.IND.
1E CONSULT BEZOEK	ECHO BEKKEN INHOUD MUV ANATOMIECODE 83 TM 86	PERITONEUM - OMENTECTOMIE
1E CONSULT POLIKLINISCH	ECHO BOVENBUIK	PIJNBESTRIJD- AANLEGGEN PATIENT CONTROLLED ANESTHESIA
339849D TOEGANGSCHIR- C.V.D.-CATHETER V.JUGULARIS - INBR.	ECHO BUIKWAND	PROGESTERON MBV RADIOISOTOPEN
AANNAME LABORATORIUMONDERZOEK	ECHO HALS --WAARONDER SCHILDKLIER--	PROLACTINE MBV EIA BLOED
ANALGESIE - EPIDURAAL DOOR ANESTHESIST	ECHO MAMMA	PUNCTIE TBV CYTOLOGISCH ONDERZOEK DOOR PATHOLOOG
ANESTHESIE - BIJ BEHANDELING MET RADIUM OF ISOTOPEN GESL	ECHO NIEREN-URINEWEGEN	PYELOGRAFIE ANTEGRAAD VIA PUNCTIE
ANESTHESIE - BIJ SPEC.ONDERZ.EN VERR.GEEN ANEST.TARIEF	ECHO ONDERBUIK	PYELOGRAFIE --WO NIERTRANSPLANTAAT--
ANTI-CHLAMYDIA IGA MBV ELISA BLOED	ECHO PANCREAS EN MILT	RESISTENTIEBEPALINGEN - 5 BEPALINGEN BACT
ANTI-HEPATITIS-C-VIRUS MBV ELISA BLOED	ECHO THORAX	RIB-STERNUM 1 R
ANTI-HIV MBV ELISA	ECHOGRAFIE - BLAAS - EXTERN	SCHOUDERGEWRICHT 2 R - LINKS
ANTI-RUBELLA IGG -ELISA - IN EEN MONSTER	ECHOGRAFIE - GENITALIA INTERNA	SCHOUDERGEWRICHT 2 R - RECHTS
BACTERIOLOGISCH ONDERZOEK MET KWEEK -NIET TBC- MYCO	ENDORETROGRADE CHOLANGIOGRAFIE	SECT.CAESAR.- ZONDER VOORBEHANDELING MET KRAAMBED
BEHANDELTIJD - EENHEID T1 - MEGAVOLT	FSH ENZYM-IMMUNOLOGISCH BLOED	SEX-HORMONE-BINDING GLOBULINE MBV RIA SERUM
BEKKEN 2 R	HAND 2 R - LINKS	SIMULATOR - GEBRUIK VOOR AANVANG MEGAVOLT BESTRALING
BEKKEN - STAAND 1 R	HART - DOPPLER ECHOCARDIOGRAFIE 2-DIMENS. M.MODE	SKELET - SCINTIGRAFIE TOTALE LICHAAM TC-99M DIFOSF
BETASUBUNIT BEP. MBV. RIA HCG DIV.MAT	HELE BOVENBEEN MET ONDERBEEN EN VOET 2 R	SPRAAKAUDIOMETRIE - STANDAARD - AUDIOLOGISCH CENTRUM
BLAAS - CYSTOSCOPIE	HEPATITIS-B SURFACE ANTIGEEN CONFIRMATIE - NEUTRAL. ELISA	SQUAMOUS CELL CARCINOMA MBV EIA BLOED-LIQUOR
BLAAS - CYSTOSCOPIE MET BIOPSIE	HEPATITIS-C-VIRUS ROCHE -HCVR- MBV PCR	T3-UPTAKE LATENTE BINDINGSCAP. GEB. TBG VOOR T3 RIA SERUM
BLAAS - SUPRAPUBISCHE KATHETER INBRENGEN	HEUP 1 R - RECHTS	TELEFONISCH CONSULT KOSTEN OUT
BLAAS - URETROCYSTOSCOPIE NNO	HISTOLOGISCH ONDERZOEK - BIOPTEN NNO	TELETHERAPIE - MEGAVOLT FOTONEN BESTRALINGSZITTING
BLAAS - URETROCYSTOSCOPIE NNO	HISTOLOGISCH ONDERZOEK - GROTE RESECTIEPREPARATEN	TESTOSTERON MBV RADIOISOTOPEN BLOED
BOVENBEEN 2 R - LINKS	HISTOLOGISCH ONDERZOEK - KLEINE RESECTIEPREPARATEN	THORAX 1 R
BRACHYTHERAPIE - INTERSTITIEEL - BIJZONDER	HOOGFREQUENTE AUDIOMETRIE - AUDIOLOGISCH CENTRUM	THORAX 2 R
BUIK - ADHESIOLYSE EN BIOPSIE DMV LAPAROSCOPIE	HUID - VERWIJDEREN RETENTIE CYSTE	THORAX - OP ZAAL 1 R
BUIK - LAPAROSCOPIE - DIAGNOSTISCH N.N.O.	HUIDPLASTIEK- Z-PLASTIEK HUID HOOFD EN HALS	TOONAUDIOMETRIE - AUDIOLOGISCH CENTRUM
BUIK - LITTEKENBREUK - PLASTISCHE OPERATIE	HYPERTHERMIE	TOTAAL T3 - TRIJODTHYRONINE MBV RADIOISOTOPEN BLOED
BUIK - PRIMAIRE STAGERINGSOPERATIE - OVARIUMCARC.	HYPERTHERMIE BEHANDELING - H-1 -	TOTAAL T4 - THYROXINE IMMUNOFLUORIMETRISCH BLOED
BUIK - PROEFLAPAROTOMIE	HYPERTHERMIE BEHANDELING - H-3 -	TSH ENZYM-IMMUNOLOGISCH BLOED
BUIK - SECOND-LOOK-OPER.OVARIUMCARC.-GEEN OMENTECT	IMMUNOPATHOLOGISCH ONDERZOEK	TUBA UTERINA- REANASTOMOSE NA STERILISATIE - ENKELZIJDIG
BUIK - STAG.LAPAROTOM.-OMENTECT.-TUMORRED.-PELVECT	INFUUS - BLOED MEER DAN 2 UUR	TUMOR - PETSCAN TOTALE LICHAAM 18-F DESOXYGL
BUIK - STAG.LAP-TUMORRED.-OMENTECT-PELVECT-LYMFECT	INFUUS - INBRENGEN	URINEWEGEN - URODYNAMISCH ONDERZOEK-VIJF OF MEER KANALEN
BUIK - STAGERINGSLAPAROTOM.-OMENTECT.-TUMORREDUCT.	INTERC.CONSULT KLINISCH ANESTHESIE	UTERUS - CARC. CERVIX VLGS. WERTHEIM-MEIGS
BUIK - STAGERINGSLAPAROTOM.-OMENTECTOMIE OVARIUMCA	INTERC.CONSULT KLINISCH CHIRURGIE	UTERUS - EXCISIE-DESTRUCTIE PATH.AFW.- HYSTEROSCOPIE
BUIKOVERZICHT 1 R	INTERC.CONSULT KLINISCH LONGZIEKTEN	UTERUS - EXTIRP.ABD.RAD. - LYMFADENECT - RE ADNEX
CERVIX - CONISATIE	INTERC.CONSULT KLINISCH RADIOTHERAPIE	UTERUS - EXTIRP.ABD.VAG.RAD.- LYMFADENECT - LI ADNEX
CERVIX - LISEXCISIE PORTIO - DIATHERMISCH	KNIE - M-Z ONDERBEEN 2 R - LINKS	UTERUS - EXTIRPATIE - ABDOMINAAL MET ADNEXA
CERVIX - LISEXCISIE PORTIO - DIATHERMISCH	Lab Klinische Chemie	UTERUS - EXTIRPATIE ABDOMINAAL RADICAAL MET ADNEXA
CITO HISTOLOGISCH ONDERZOEK	LH MBV RADIOISOTOPEN BLOED	UTERUS - EXTIRPATIE ABDOMINAAL TOTAAL
CLOSTRIDIUM ELISA-TEST	LIGASE CHAIN REACTION -LCR-	UTERUS - EXTIRPATIE VAGINAAL TOTAAL
COLON INLOOP	LIGDAGEN - ALLE SPEC.BEH.KINDERG.-REVAL.-RADIOTH.-ANESTH.	UTERUS - HYSTEROSCOPIE
COUPE TER INZAGE	LONGFUNCTIE - CO2 - CAPNOGRAFIE	UTERUS - INTRA-UTERINE DEVICE INBRENGEN-ANTICONCEPTI
CT A.PULMONALIS M C	LYMF.SYST. - BIOPSIE LYMFEKLIER	VAGINA - BIOPSIE-PUNCTIE-CYTOLOGIE
CT ABDOMEN M C	LYMF.SYST. - REGIONALE LYMFEKLIERDISSECTIE - ENKELZIJDIG	VAGINA - COLPOCLEISIS LEFORT
CT BEKKEN M C	LYMF.SYST. - SENTINEL NODE PROCEDURE LIES	VAGINA - OPERATIE VESICOVAGINALE FISTEL
CT BOVENBUIK M C	LYMF.SYST. - STAGERINGSLYMFADENECTOMIE KLEINE BEKKEN	VAGINA - OVERIGE EXCISIE PATHOLOGISCHE AFWIJKINGEN
CT HALS M C	LYMFEKLIER - SCINT SENTINEL NODE DYNAMISCH TC-99M COLLOI	VAGINA - SCOPIE EXCL.WEEFSELWEGNAME- SPECULUMONDERZ.
CT HERSENEN M C	LYMFEKLIER - SCINT SENTINEL NODE MET PROBE TC-99M COLLOI	VAGINA - SCOPIE INCL.EVT.VULVABIOPS.NIET MET HYST.SC
CT LEVER EN GALWEGEN M C	LYMFEKLIER - SCINT SENTINEL NODE VERVOLG TC-99M COLLOI	VAGINA - TENSION-FREE VAGINAL TAPE IVM INCONTINENTIE
CT THORAX M K	MAAG - ONTLEDIG.-SCINT VAST-VLOEIB.VOEDSEL IN-111 TC-99M	VAGINA - TOUCHER ONDER ANESTHESIE
CT THORAX Z K	MAMMOGRAFIE THORAXWAND	VERVOLGCONSULT POLIKLINISCH RTH DECLARABEL
CT TOTALE LICHAAM	MANDIBULA-KAAKGEWRICHT 1 R - BDZ	VRIESCOUPE
CT-GEL. ALS ASS. BIJ PUNCTIE - BIOPSIE INTRAPERITONEAAL	MEC-A-BEPALING VOOR MRSA MBV PCR	VRIJ T4 MBV RIA -FT 4 BLOED
CYSTOGRAFIE RETROGRAAD	MICROSCOPISCH ONDERZOEK - ELEKTRONENMICROSCOOP -	VRW.GESL.ORG- ADNEX-EXTIRPATIE DMV LAPAROSCOPIE-DUBBELZ.
CYTOLOGISCH ONDERZOEK - ASCITES -	MICROSCOPISCH ONDERZOEK - GEKLEURD EN ONGEKLEURD	VRW.GESL.ORG- ADNEX-EXTIRPATIE DMV LAPAROTOMIE-DUBBELZ.
CYTOLOGISCH ONDERZOEK - BUIKTUMORPUNCTIE -	MORFOMETRIE	VRW.GESL.ORG- ADNEX-EXTIRPATIE DMV LAPAROTOMIE-ENKELZ.
CYTOLOGISCH ONDERZOEK - DIVERSEN -	MRI ABDOMEN	VRW.GESL.ORG- BIOPSIE-PUNCTIE-CYTOLOGIE CERVIX-UTERUS
CYTOLOGISCH ONDERZOEK - ECTOCERVIX -	MRI BEKKEN	VRW.GESL.ORG- CURETTAGE GEFRACTIONEERD
CYTOLOGISCH ONDERZOEK - LEVERPUNCTIE -	MRI BIJNIER	VRW.GESL.ORG- EXC.DESTRUCT.PATH.AFWIJKING VULVA-PERINEUM
CYTOLOGISCH ONDERZOEK - LYMFEKLIERPUNCTIE -	MRI WERVELKOLOM - LUMBAAL	VRW.GESL.ORG- OPERATIE VAN CLITORIS NNO
CYTOLOGISCH ONDERZOEK - PLEURAVOCHT -	MRI WERVELKOLOM - THORACAAL	VRW.GESL.ORG- PLASTISCHE OPERATIE VAN VULVA OF PERINEUM
CYTOLOGISCH ONDERZOEK - URINE NIERCYSTE URETERURINE -	MYCOBACTERIUM MBV PCR DIV.MAT	VRW.GESL.ORG- SECOND-LOOK-OPER.OVARIUMCARC.LAPAROSCOPISCH
CYTOLOGISCH ONDERZOEK - VAGINA -	NIET DECLARABELE DAGVERPLEGING-BV KLIN.PAT.ELDERS	VULVA - BIOPSIE-PUNCTIE-CYTOLOGIE
CYTOLOGISCH ONDERZOEK TBV BEVOLKINGSONDERZOEK	NUCLEAIR ONDERZOEK VAN ELDERS TELCODE	VULVA - INCISIE - OVERIGE
DAGVERPLEGING - ALLE SPEC.BEH.KIND.-REV.-RADIOTH.-ANESTH.	OBDUCTIE	VULVA - PARTIELE VULVECTOMIE
DARM - COLONPASSAGE SCINT.COMP.VERVOLG TC-99M COLLOI	OEFENTHERAPIE - BEWEGINGSTHERAPIE	VULVA - RAD.VULVECTOMIE - OPP.EN DIEPE LIESKL.DISS.
DELTA-4-ANDROSTEENDION BLOED	OESTRADIOL MBV RADIOISOTOPEN -RIA BLOED	VULVA - RADICALE VULVECT. ZND INGUINALE LYMFADENECT
DIAGNOSTISCHE PUNCTIE ONDER ECHOCONTROLE	OESTRON MBV RADIOISOTOPEN BLOED	VULVA - RUIME LOKALE EXCISIE VAN AANDOENING
DIEET NNO	ONDERZOEK VAN ELDERS NIET BEOORDEELD	VULVA - VULVECTOMIE - LIESBLOK
DIKKE DARM - APPENDECTOMIE NNO	OPNAME VOLGENS BARSONIE 1 R	VULVA - VULVECTOMIE ZONDER LIESKLIEREXTIRPATIE
DIKKE DARM - RESECTIE SIGMOID MET PRIMAIRE ANASTOMOSE	OVARIUM - ADNEX-EXTIRPATIE DMV LAPAROSCOPIE	WERVELKOLOM - LUMBAAL 2 R
DUPLEXSCAN - VENEN BEEN	OVARIUM - DEBULKING OVARIUMCARCINOOM	
ECHO ABDOMEN	OVARIUM - OVARIOPEXIE -EXCL.TORSIE-	

Table 11: List of the activities included in the analyses after filtering the administrative activities

11. APPENDIX C

Table 12 to Table 18 present all the activities included in the method of Riemers 2009 for the different phases of the method. Thus, the building database phase is illustrated in Table 12, the introduction session phase in Table 13, and so on. The last column of the tables identifies if each individual activity was performed during the validation process of this graduation project.

	Contact data manager	✗
Extract data	Define business goal	✓
	Identify the data warehouse	✓
	Start data collection	✓
	Integration and consolidation of the data	✗
Transform data	Convert to MVN-format	✓
	Pre-analysis and pre-processing of the data	✓
	Create database and import data	✓
Load	Use ProM import tool	✗
	Load data into MagnaView	✓

Table 12: Comparison between the actions included in the building database phase of the method and the actions done during this project.

Meet specialists	✓
Determine users	✓
Determine status of change	✓
Communicate / determine goals	✓
Determine first set of KPIs and extra	✓
Obtain information about HC program (domain knowledge)	✗

Table 13: Comparison between the actions included in the introduction session phase of the method and the actions done during this project.

	All activities	✓
	First activities	✓
	Final activities	✓
	Patterns (sequential)	✓
	Patterns (grouped)	✓
	Process mining results	✓
	Centered activity with patterns (sequential)	✗
	Centered activity with patterns (grouped)	✗
	Centered activity with causal relations (1 step)	✗

	Centered activity with causal relations (3 steps)	✗
KPIs	General information	✓
	Overview	✓
	Results	✓
	Validation (process analysis)	✓

Table 14: Comparison between the actions included in the preliminary analysis phase of the method and the actions done during this project.

Introduction	Explain (calculation) method	✓
	Discuss initial results	✓
	Let actor(s) interactively "play" with tool	✗
Selection	Select activities that do not belong in treatment process	✓
	Select activities that are mandatory in treatment process	✓
	Determine outliers	✓
Closure	Determine follow-up steps	✓
	Extra KPI's	✓
	Focus process visualization / analysis	✓

Table 15: Comparison between the actions included in preliminary meeting phase of the method and the actions done during this project.

	All activities	✓
	First activities	✗
	Final activities	✗
	Patterns (sequential)	✓
	Patterns (grouped)	✓
	Process mining results	✓
	Centered activity with patterns (sequential)	✓
	Centered activity with patterns (grouped)	✗
	Centered activity with causal relations (1 step)	✗
	Centered activity with causal relations (3 steps)	✗
KPIs	General information	✓
	Overview	✓
	Results	✓
	Validation (process analysis)	✓

Table 16: Comparison between the actions included in the 2nd analysis phase of the method and the actions done during this project.

Present & discuss results	✓
Receive feedback on results	✓
Determine which goals were reached	✓
Determine follow-up steps	✓

Table 17: Comparison between the actions included in the final meeting phase of the method and the actions done during this project.

Project document	Summary		✓
	General information	Goals	✓
		Target group	✓
		Data	✓
	List of all activities		✓
	List of first activities		✗
	List of final activities		✗
	Most important patterns		✓
	Most important process mining results		✓
	Most important centered view		✓
	KPIs	General information	✓
		Overview	✓
		Results	✓
		Validation	✓
	Conclusions		✓
Project file MagnaView			✓

Table 18: Comparison between the actions included in the final document of the method and the actions done during this project.

12. APPENDIX D

Table 19 presents a list of how the renaming in the preliminary analysis phase was done. As is shown, the table shows two columns. The first column includes the original names of the activities. Every activity that has this name at the activity or department level was renamed with the name in its right.

ORIGINAL NAME	RENAME
Algemeen Lab Klinische Chemie	Lab
Radiologie	Radiology
Operatiekamers	Operating Rooms
Pathologie	Pathology
IC Volwassenen	IC
Nucleaire Geneeskunde	Nuclear Healthcare
Specieel Lab Endo/Radio	Lab Endo/Radio
Vaatlaboratorium	Vascular Lab
Verkoever/High Care	Recovery Room_High Care
Functie centrum	Function Center ENT
Endoscopie	Endoscopy
Dietetiek	Dietics
CT ABDOM.MC 387042A	Radiology
Hart- en Vaat poli	Heart and Vascular Diseases Outpatient Clinic
Kraamafd. H3Z EN H4Z	Maternity Ward H3Z and H4Z
Verpleegafdeling F6Z	Nursing Ward F6Z
Poli Anesthesiologie	OC Anaesthetics
Radiotherapie	Radiotherapy
Verpleegafdeling F7N	Nursing Ward F7N
Verpleegafdeling F7Z	Nursing Ward F7Z
Verpleegafdeling G5Z	Nursing Ward G5Z
Dagcentrum - verpleegafdeling	Daycare
H5NO URO	Nursing Ward H5N
Verpleegafdeling F3 Noord	Nursing Ward F3N
F5N	Nursing Ward F5N
Verpleegafdeling F6N	Nursing Ward F6N
Medische Microbiologie	Medical Microbiology
Verpl.afd. G5NZ	Nursing Ward G5NZ
Verpl.afd. G6NO/G6ZU	Nursing Ward G6N/G6Z
Lab bepalingen door derden	Lab External
Kliniek H8Z Tieners	Nursing Ward H8Z
Verpleegafdeling F3 Zuid	Nursing Ward F3Z
Verpleegafdeling H7N/Z NEU	Nursing Ward H7
BEZOEK Dagcentrum - pijnbestrijding	Daycare Pain Control
BEZOEK Poli CHI/URO	OC Surgery/Urology
Poli neurologie	OC Neurology
Poli Inwendige Specialismen	OC Inward Specialties
Poli CAP/CHP/ORT/TRA A1-4	OC CAP/CHP/ORT/TRA
Dagcentrum - behandelcentrum	Daycare
Lab. Exp. Immunologie	Lab Immunology
GYNAECOLOGIE H5Z	Nursing Ward H5Z
Poli IVF en ENVO	OC IVF
1E CONSULT BEZOEK Polikliniek Verlosk.-Gyn.	OC Gyn Onc
1E CONSULT POLIKLINISCH Polikliniek Verlosk.-Gyn.	OC Gyn Onc
Poli CHI/URO	OC Surgery/Urology
EERSTE POL Polikliniek Verlosk.-Gyn.	OC Gyn Onc
Spoedeisende Hulp	Urgent Care
Specieel Lab Neurozintuigen	Lab Neurosenses
Hyperpressietank	Hypercare
Klinische neurofysiologie	Clinical neurofysiology
Spec.Lab. Hematologie	Lab Hematology
Verpleegafdeling H6ZU NEC	Nursing Ward H6Z
Longfunktie Onderzoek	Lung Diseases
CARDIOVASCULAIRE STENT Hartcatheterisatie	Cardiology
Poli Oogziekten	OC Eye Diseases
Polikliniek Verlosk.-Gyn.	OC Gyn Onc
Specieel Lab GMZ	Lab GMZ
Poli pijnbestrijding	OC Pain Control
INTERNE ZIEKTEN	Internal Diseases
Apotheek Laboratorium	Pharmacy Lab
KLINISCH INTERNE	Internal Diseases
Dagcentrum - pijnbestrijding	Daycare
Poli MZK A1-1	OC MZK

Table 19: Rename of activities

13. APPENDIX E

According to Song et al. 2008, the Self-Organizing Map (SOM) plug-in is a neural network technique that groups similar cases close together in certain areas of a value range. The usage of this plug-in is recommended in the method developed by Riemers (2009). Figure 31 shows that the patients were grouped into two different groups, one of 69 patients and the other one of 613. Furthermore, the settings used to get this picture were the default settings as is shown in Figure 30.

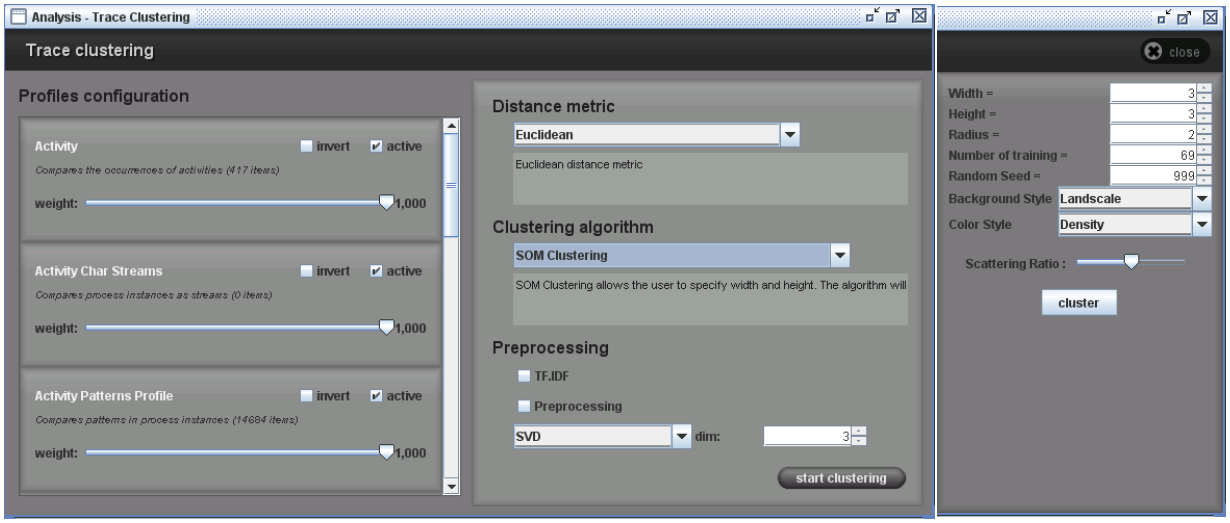


Figure 30: Trace clustering settings to get the clustering results in the preliminary analysis phase

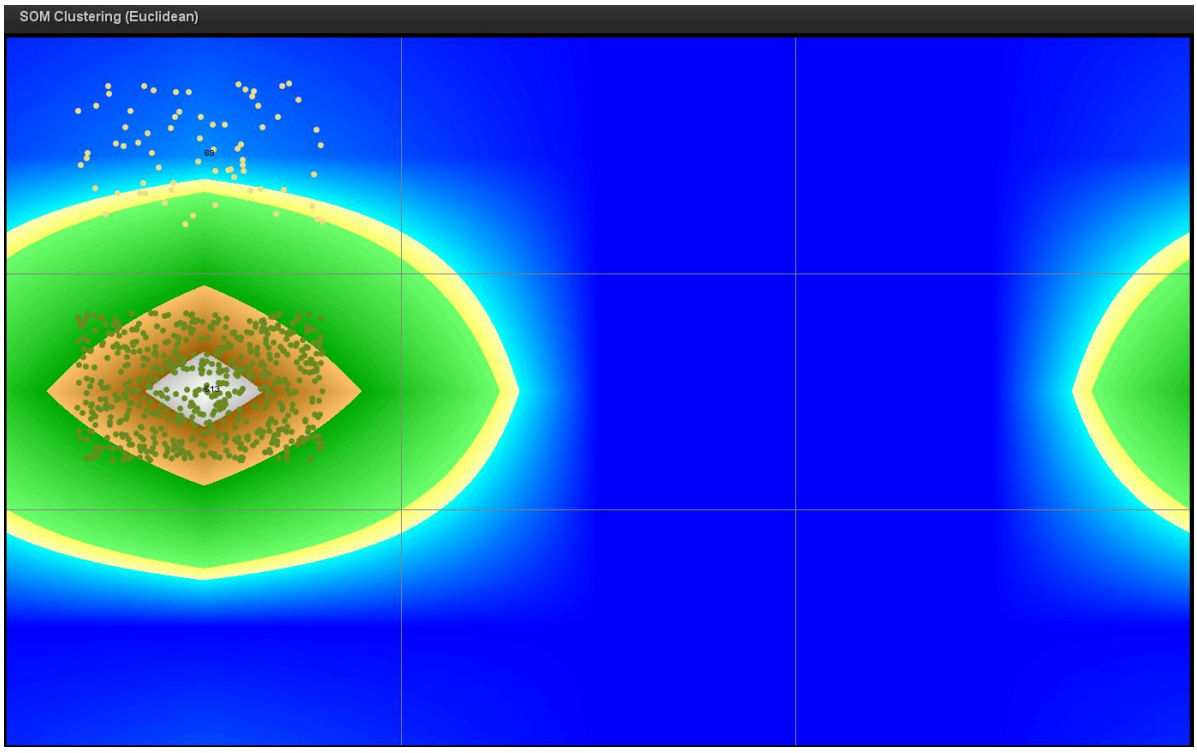


Figure 31: Clustering results on the aggregated dataset - Preliminary analysis

14. APPENDIX F

Figure 34 and Figure 35 present the process models for the groups of 69 and 613 patients respectively. These models come from the groups of patients produced by the *SOM* plug-in in the preliminary analysis phase of the validation process. Additionally, Figure 32 and Figure 33 present the settings of the *Heuristic miner* plug-in for the groups of 69 and 613 patients respectively. This mining plug-in is recommended in the method developed by Riemers (2009) as the plug-in that should be used to obtain process models.

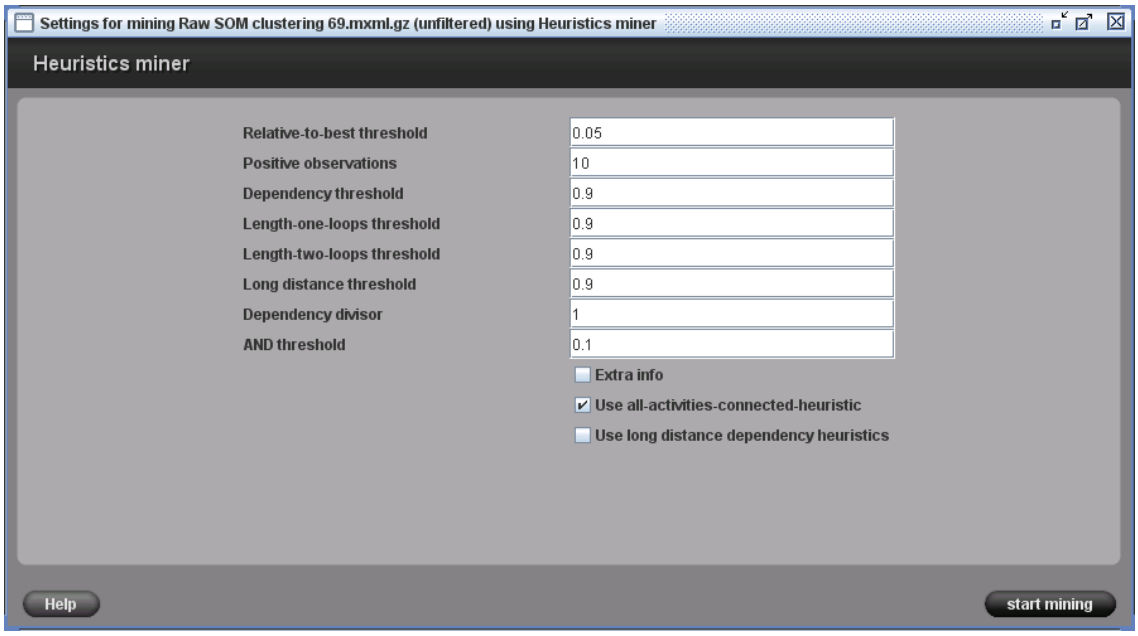


Figure 32: Settings of the Heuristic miner plug-in for the group of 69 patients

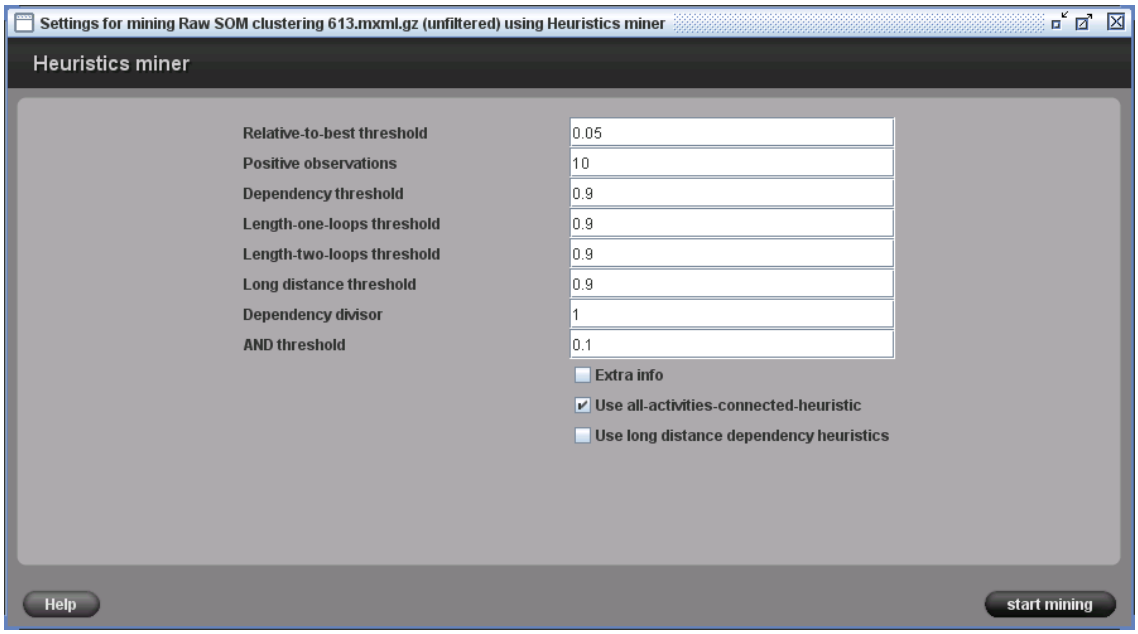


Figure 33: Settings of the Heuristic miner plug-in for the group of 613 patients

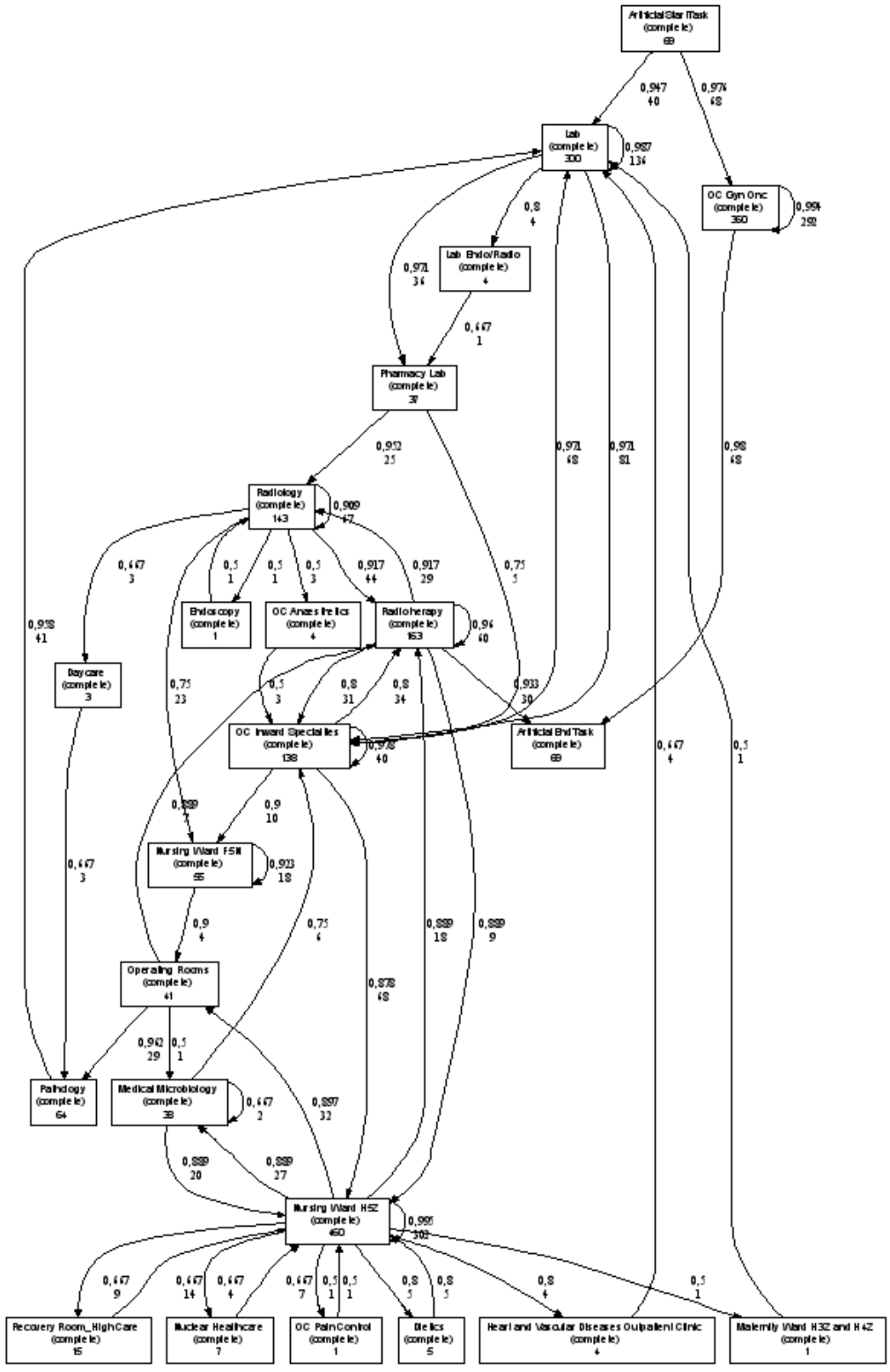


Figure 34: Process model for the group of 69 patients - Preliminary analysis

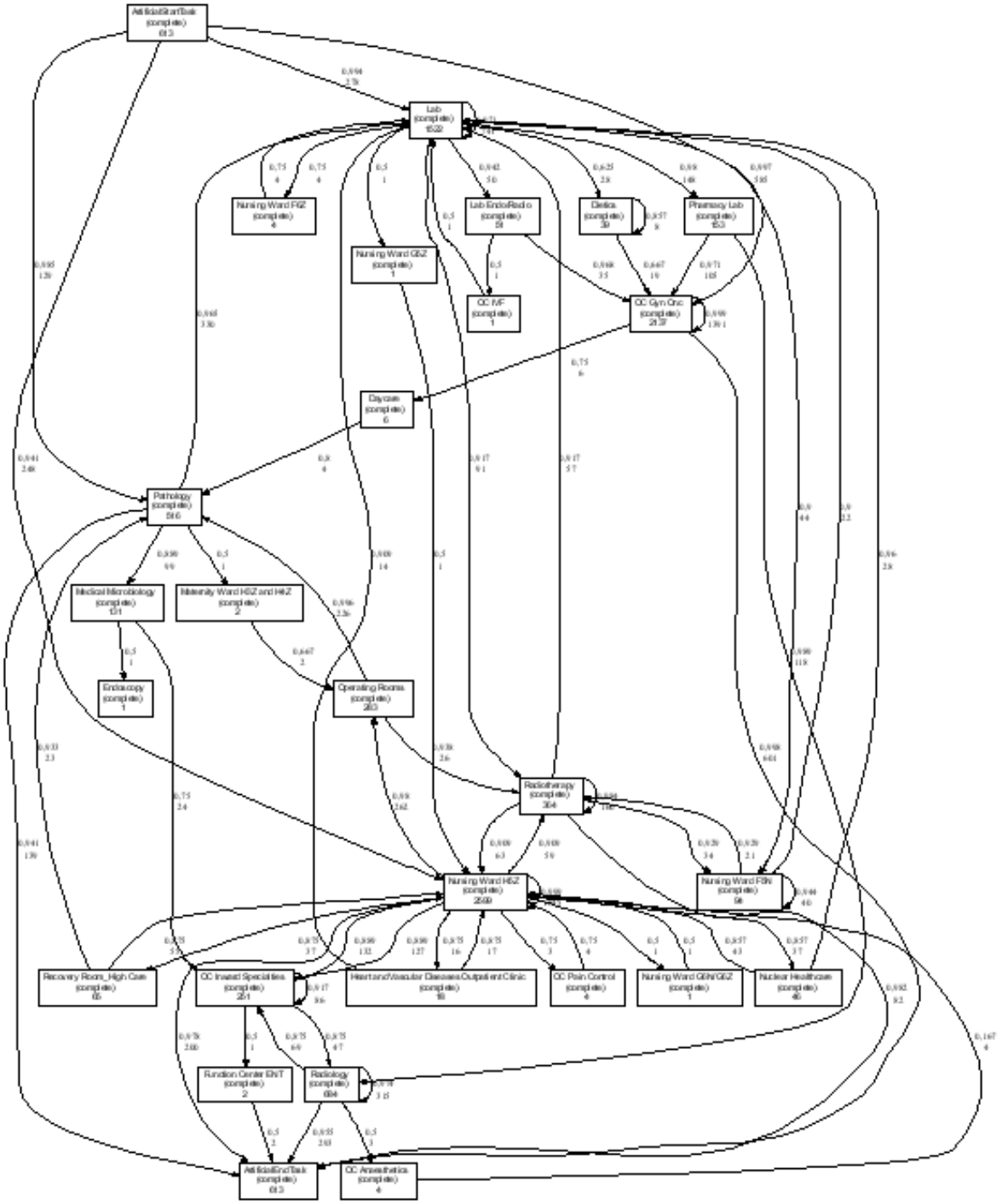


Figure 35: Process model for the group of 613 patients - Preliminary analysis

15. APPENDIX G

Figure 36 shows the results for the KPI of resource utilization as established in the preliminary analysis phase (Section 3.3). Additionally, the resulted visualization of the third KPI established also in this phase is shown in Figure 37. This last KPI of the preliminary analysis phase is related with the number of doctors seen by patients. Figure 37 presents that 315 patients were seen by one doctor, 201 by two doctors and so on. The maximum number of doctors seen was present in three patients who saw seven doctors during their process in the hospital. In the 2nd analysis phase, this analysis was repeated because the number of appointments was reduced only to the appointments in the gynecological oncology department.

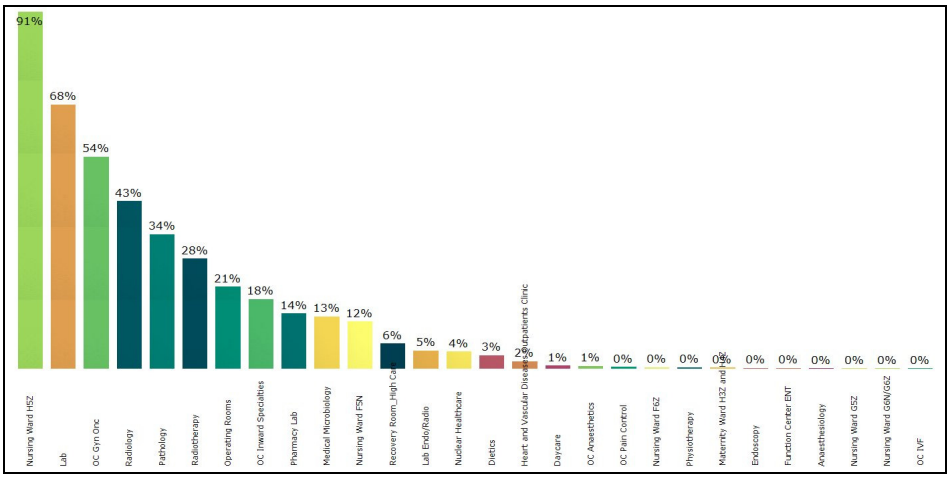


Figure 36: Resource utilization

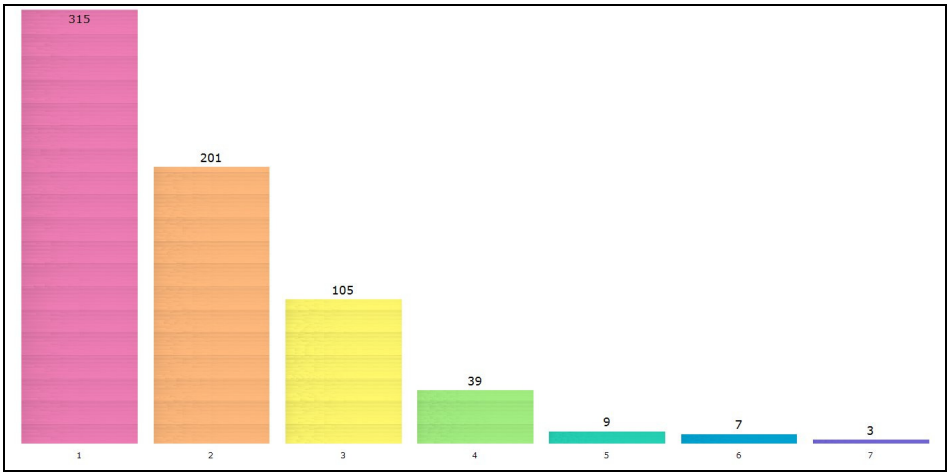


Figure 37: Number of doctors seen by patient – Preliminary analysis

16. APPENDIX H

Next paragraph presents all the specific questions generated during the preliminary meeting for each of the four main topics investigated during the validation process of this graduation project.

- Level of standardization in the clinical pathways:
    - How is the visualization of the processes?
    - Which are the most frequent activities?
    - Which is the most followed route of patients?
    - Which are the most common surgeries?
    - Which are the most common radiotherapies?
    - Are there any differences in treatment according to the age of the patients?
    - Are there any patterns before and after certain surgeries?
- The collaboration between departments:
    - Which are the most interesting handovers of work between departments?
- Logistic insights:
    - What is the average time for surgery?
    - What is the average time for surgery to radiotherapy?
    - What is the average number of appointments for patients?
    - What is the average time between the first and the second date of the patients?
- The compliance of a policy in the department which states that the patients have to be seen always by the same doctor
    - How many different doctors do the patients see when visiting the hospital?
    - What is the average number of activities per patient according to the number of different doctors seen?
    - How does the comparison of patient processes between patients that saw only one doctor and the ones that saw more than one doctor look like?

17. APPENDIX I

Before obtaining the clusters of patients with the *SOM* plug-in, some analyses were carried out in order to determine the best settings for this plug-in. Mainly, the analyses were done by changing the "Distance metric" option of this plug-in. The available metrics in this option are: Euclidean, Jaccard Index, Hamming, Correlation Coefficient, Levenshtein Edit Distance, and Generic Edit Distance. The results produced only one group of patients in all the metrics except for the Euclidean option. Therefore, this was the option chosen to the following analyses. Figure 38 presents the resulted visualization with only one group of patients with the Hamming metric in the *SOM* plug-in. As has been said, all the metrics produced the same visualization except for the Euclidean metric.

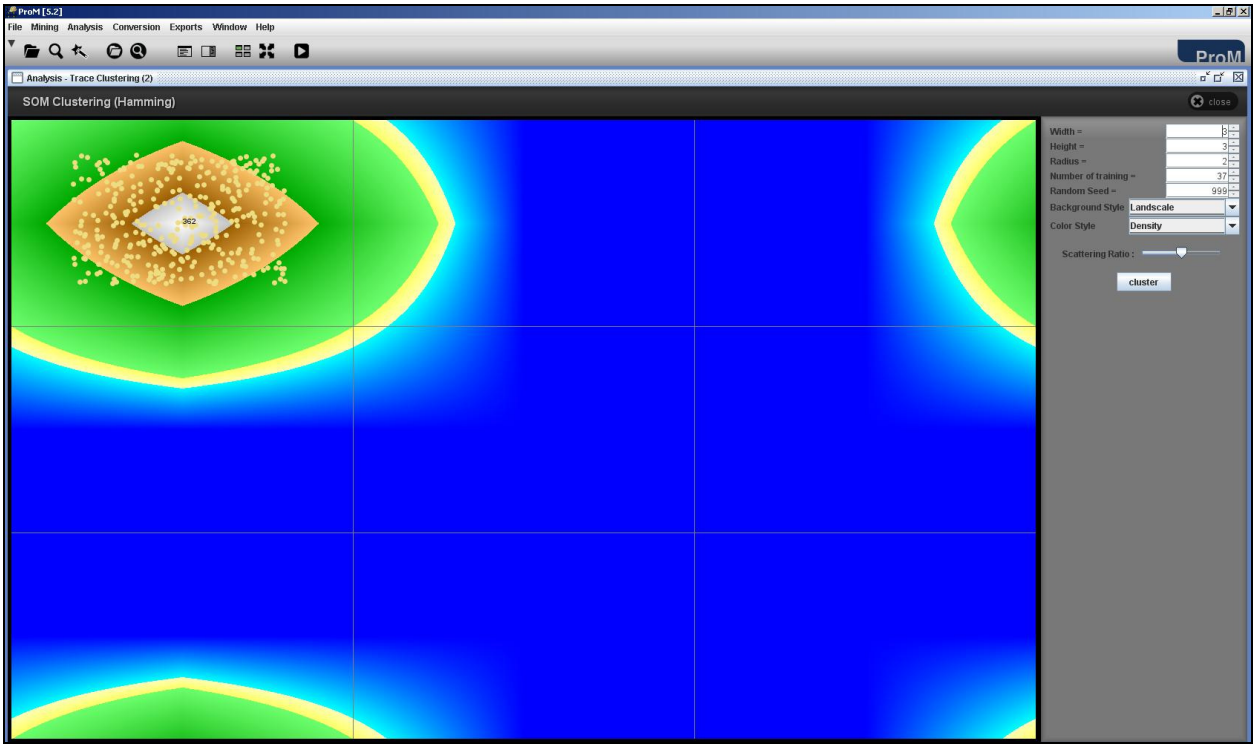


Figure 38: Clustering results with the SOM plug-in and the Hamming distance metric

Figure 39 shows the settings used for the *SOM* plug-in to get the different clusters of patients in the 2nd analysis phase. Furthermore, Figure 40 presents the visualization of the *SOM* plug-in where it is possible to see the two different groups of patients that were cluster together, one of 47 and the other one of 315 patients.

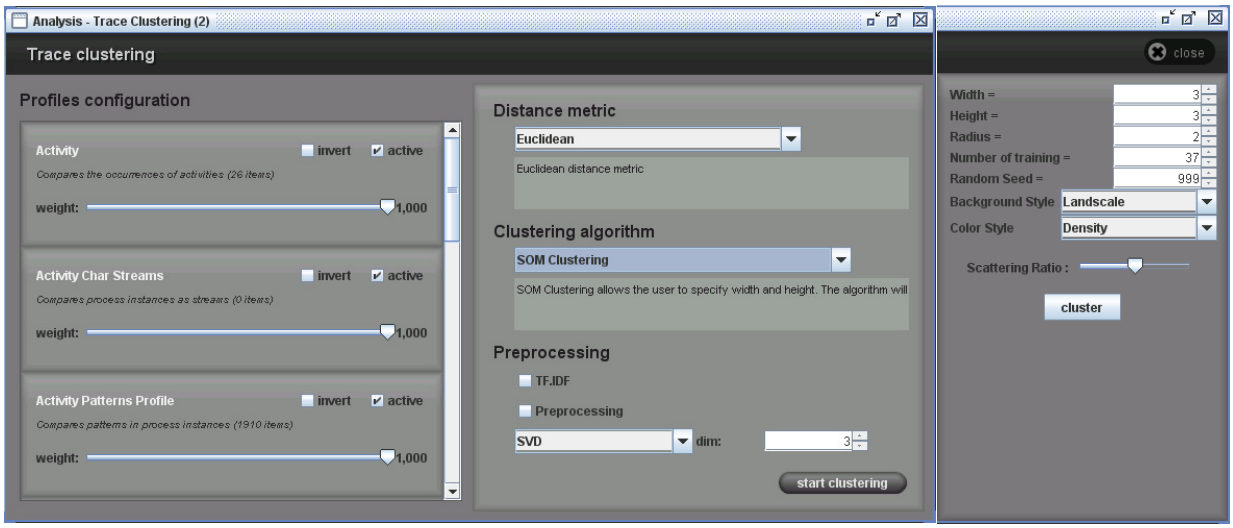


Figure 39: Trace clustering settings to get the clustering results in the 2nd analysis phase

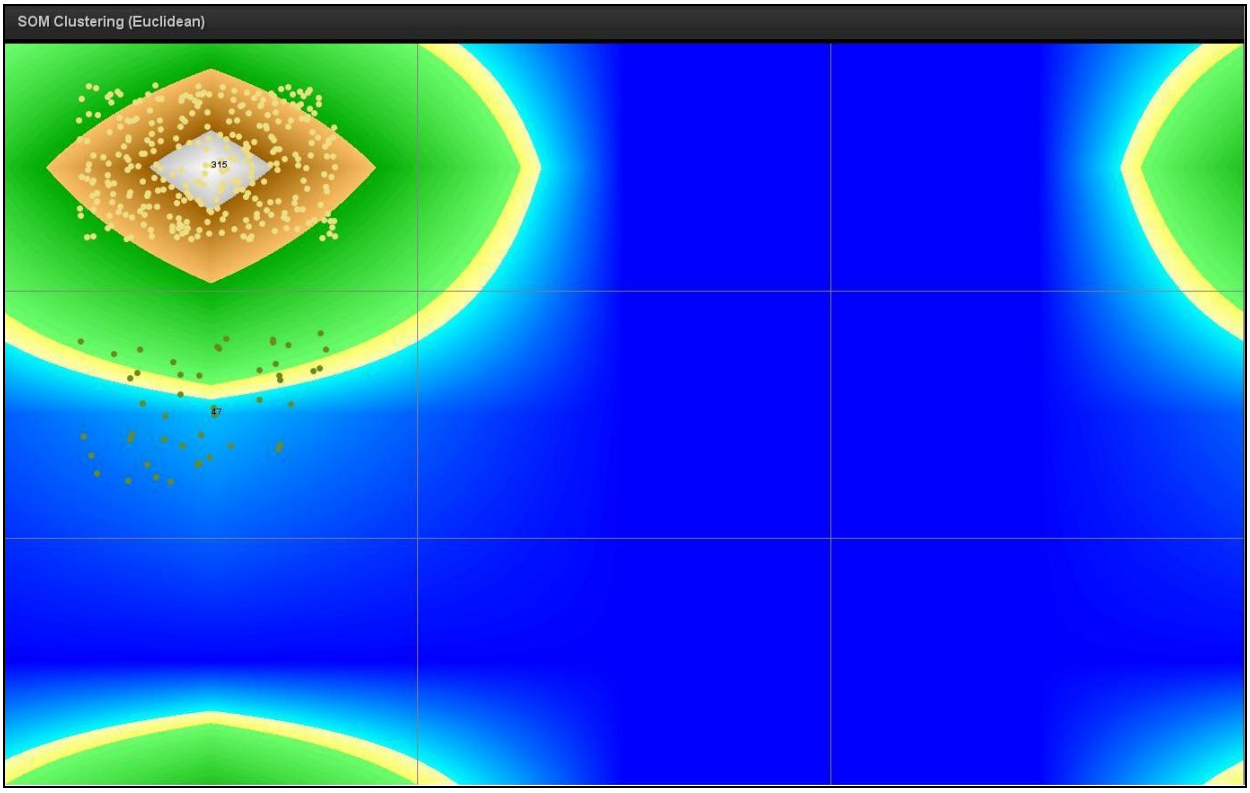


Figure 40: Clustering results on the aggregated dataset - 2nd analysis

18. APPENDIX J

Figure 43 and Figure 44 present the process models for the groups of 315 and 47 patients respectively. These models come from the groups of patients produced by the *SOM* plug-in in the 2nd analysis phase of the validation process. Additionally, Figure 41 and Figure 42 present the settings of the *Heuristic miner* plug-in for the groups of 47 and 315 patients respectively.

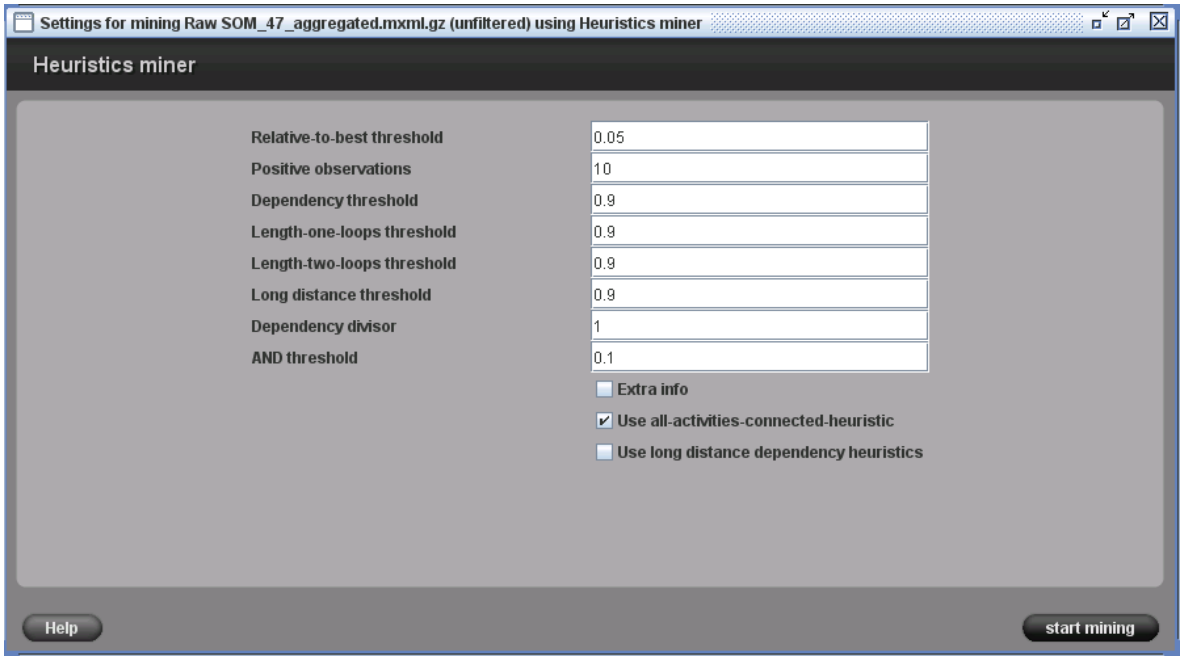


Figure 41: Settings of the Heuristic miner plug-in for the group of 47 patients

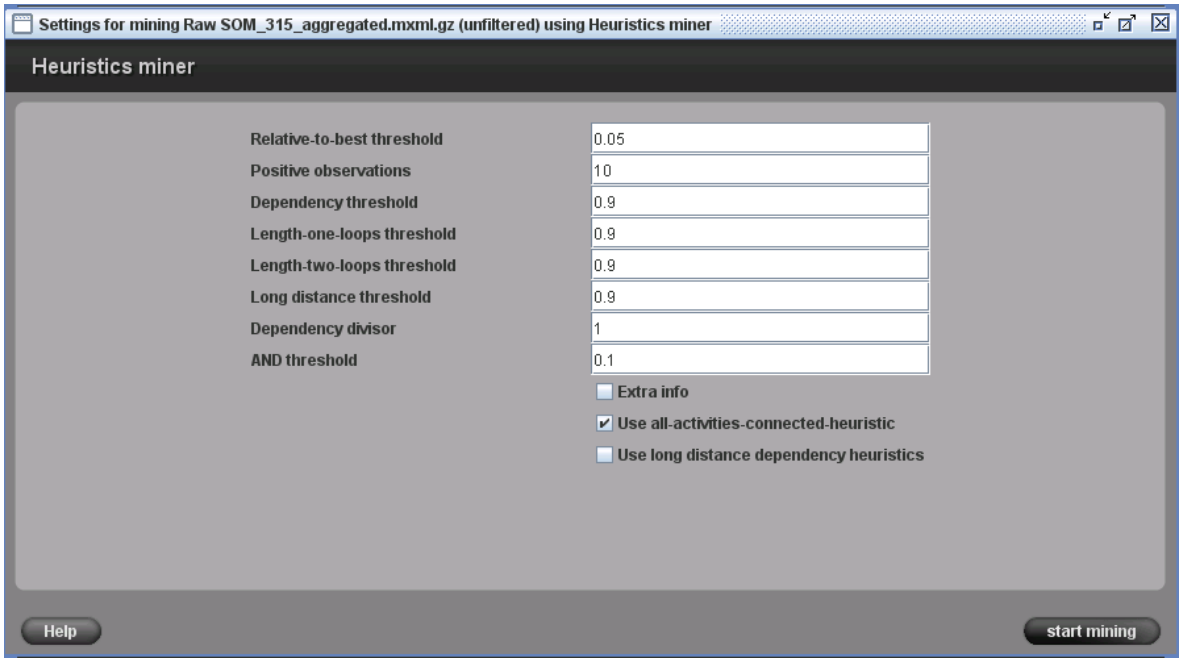


Figure 42: Settings of the Heuristic miner plug-in for the group of 315 patients

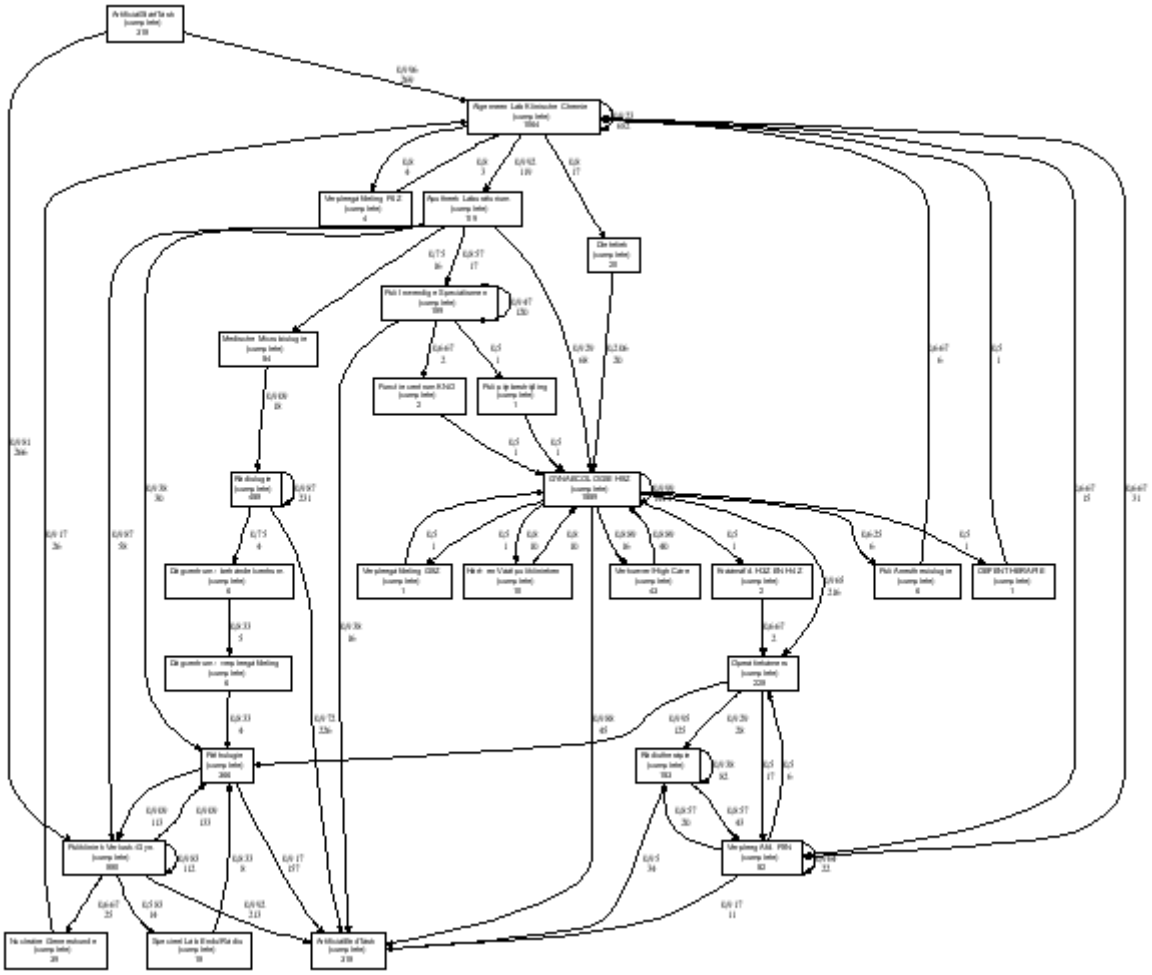


Figure 43: Process model for the group of 315 patients – 2nd analysis

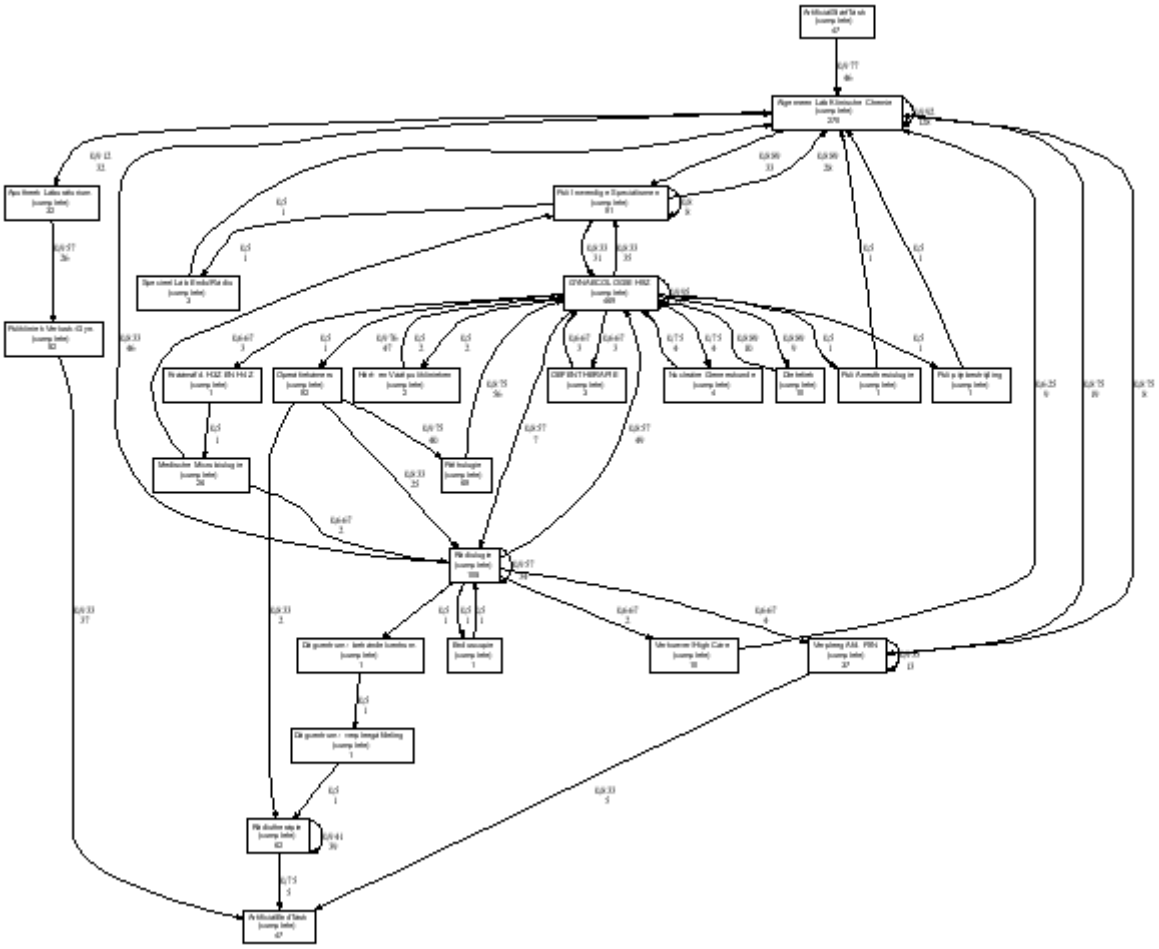


Figure 44: Process model for the group of 47 patients – 2nd analysis

19. APPENDIX K

Figure 45 shows the results for the number of doctors seen by patients. This visualization was developed in the 2nd analysis phase of the validation process and it is explained in Section 3.5.

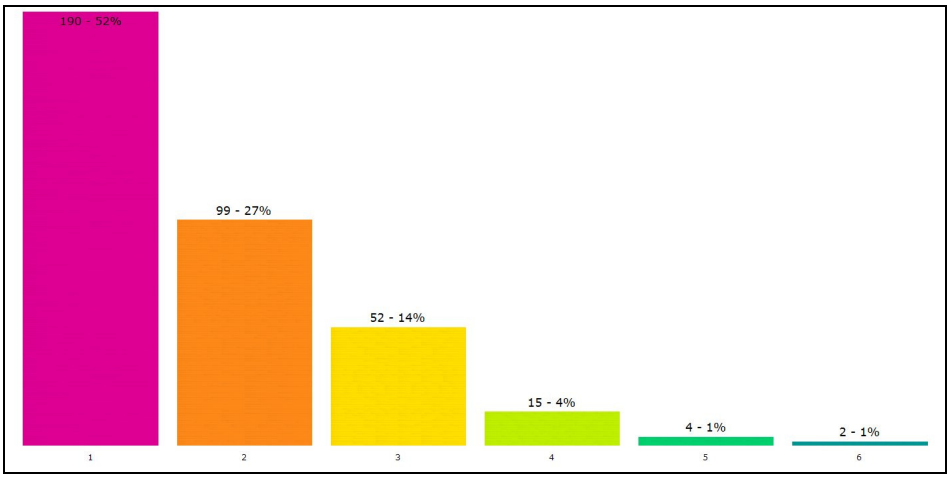


Figure 45: Number of doctors seen by patient - 2nd analysis

20. APPENDIX L

The default settings of the *SOM* plug-in were used to get the clustering results for the six different DBC-codes as explained in section 5.1.4. The distance metric used in the *SOM* plug-in was the *Euclidean*. This option has previously shown its usefulness. Figure 46 shows the visualization produced for the data in the M11 DBC-code. As is shown in the figure no clustering groups were formed and the original 50 patients remain in one group.

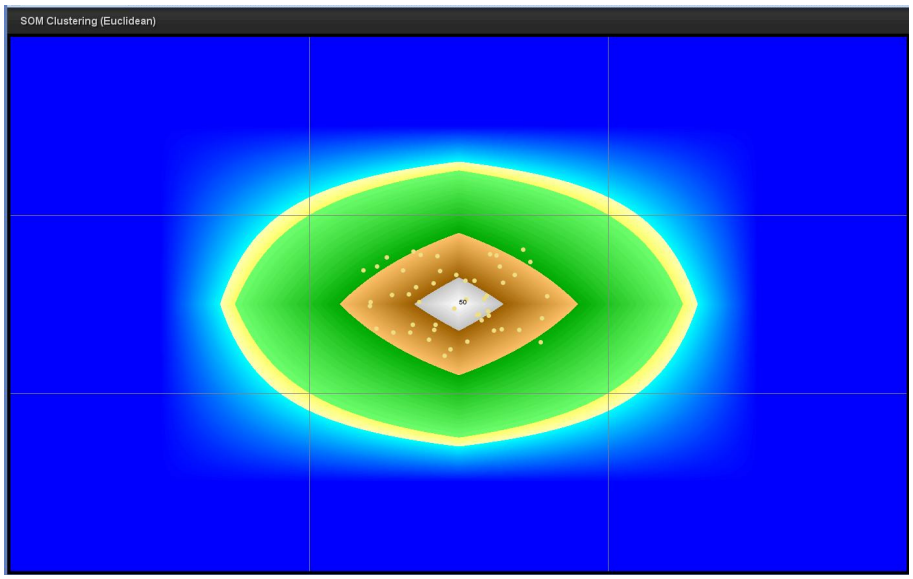


Figure 46: Clustering results for M11

Figure 47 shows the clustering visualization for M12. This DBC-code only has 5 patients and the plug-in found 3 different groups of patients, 2 groups of 2 patients and 1 group with one patient.

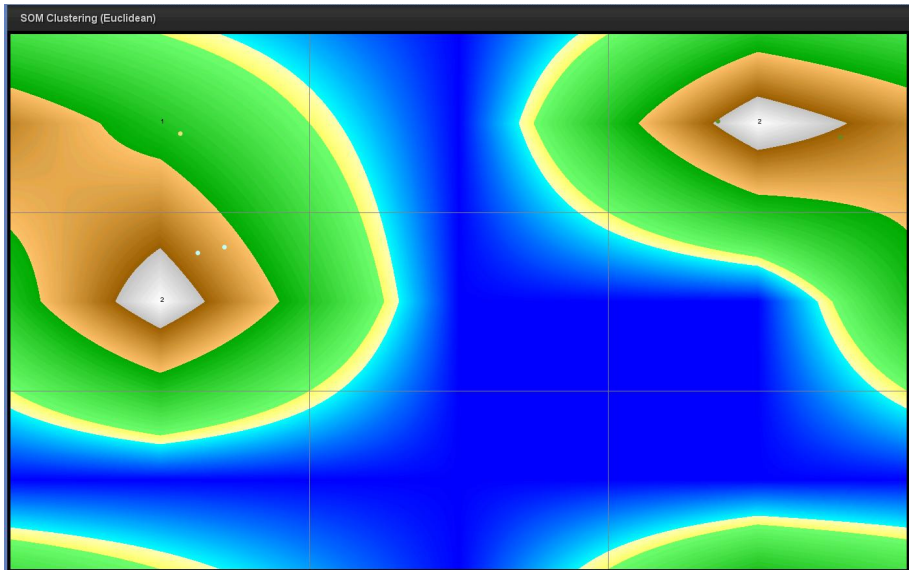


Figure 47: Clustering results for M12

Figure 48 presents the group formed after analyzing M13. As is shown the patients were cluster into one group of 138 patients so no division was made.

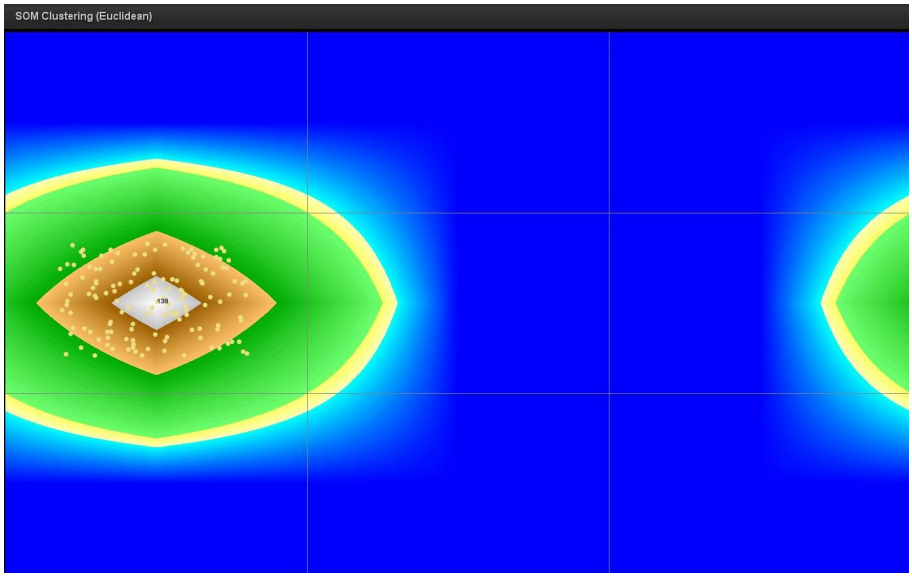


Figure 48: Clustering results for M13

Figure 49 shows the results for DBC-code M14. As is presented in this figure the patients in this DBC-code were clustered into 4 different groups. The number of patients for each group is: 53, 10, 3 and 2 patients

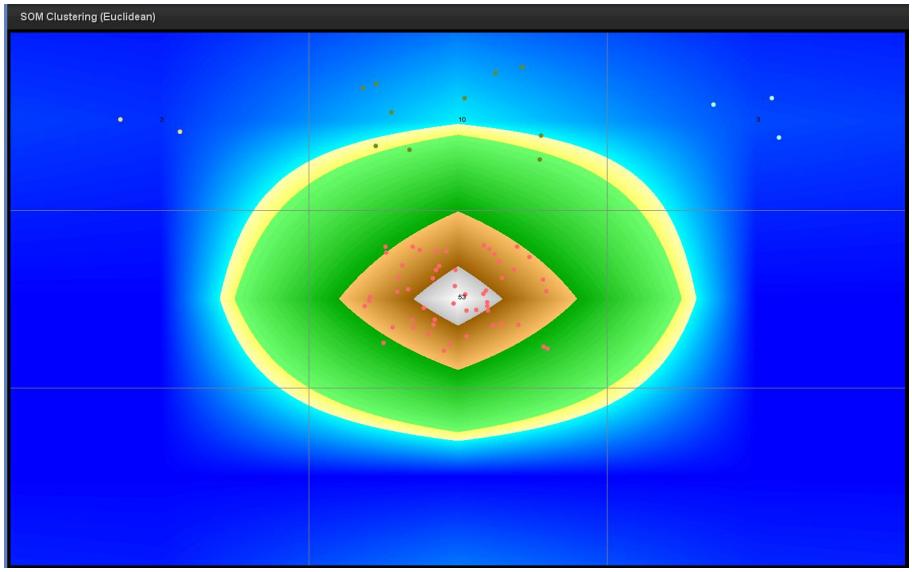


Figure 49: Clustering results for M14

Figure 50 illustrates the visualization obtained for the DBC-code M15. Three groups of patients were clustered: 3, 2 and 1 patient for each specific group.

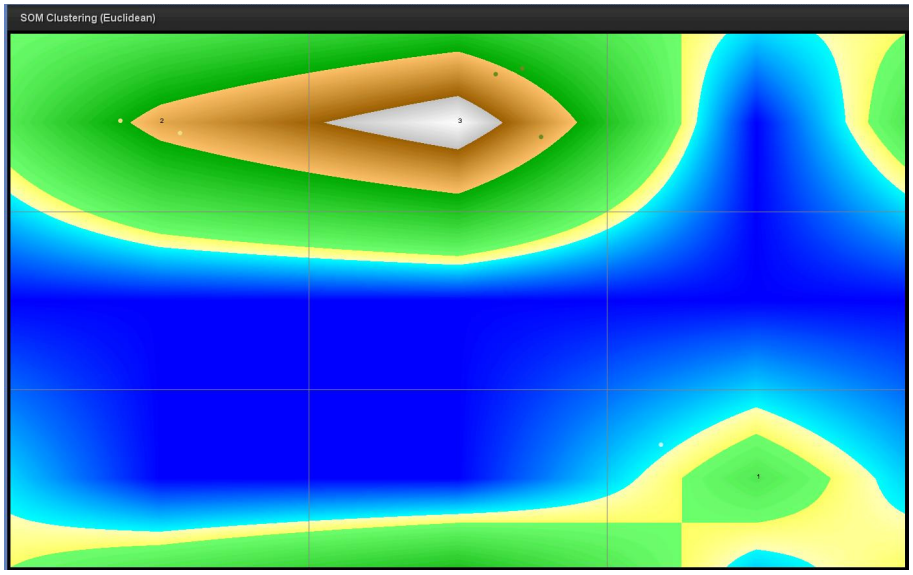


Figure 50: Clustering results for M15

And finally, Figure 51 shows the results for DBC-code M16. The groups clustered were: one group of 57 patients and two groups with 19 patients.

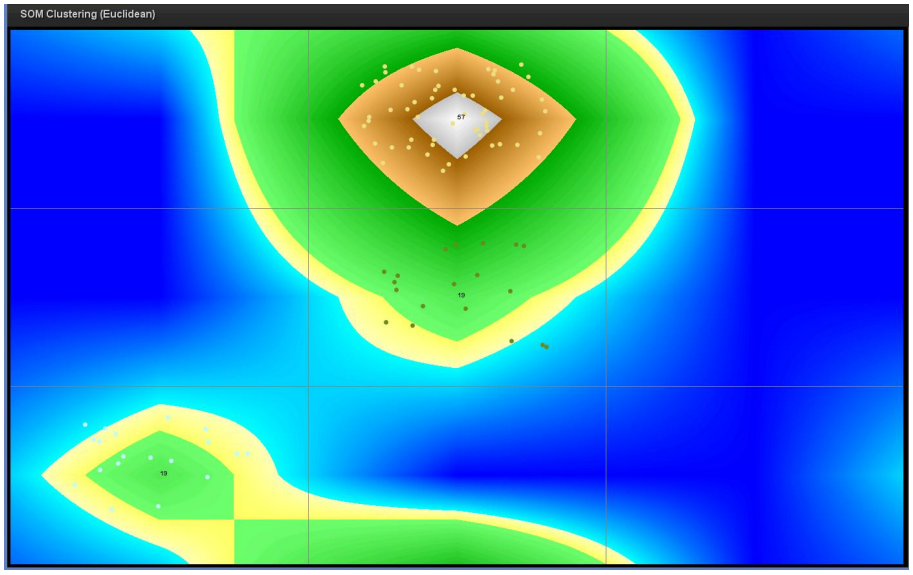


Figure 51: Clustering results for M16

21. APPENDIX M

Figure 52, Figure 53 and Figure 54 present the resulting process models for the formed group of 57 patients produced by the *Heuristic miner* plug-in, *Genetic miner* plug-in and *Fuzzy miner* plug-in respectively.

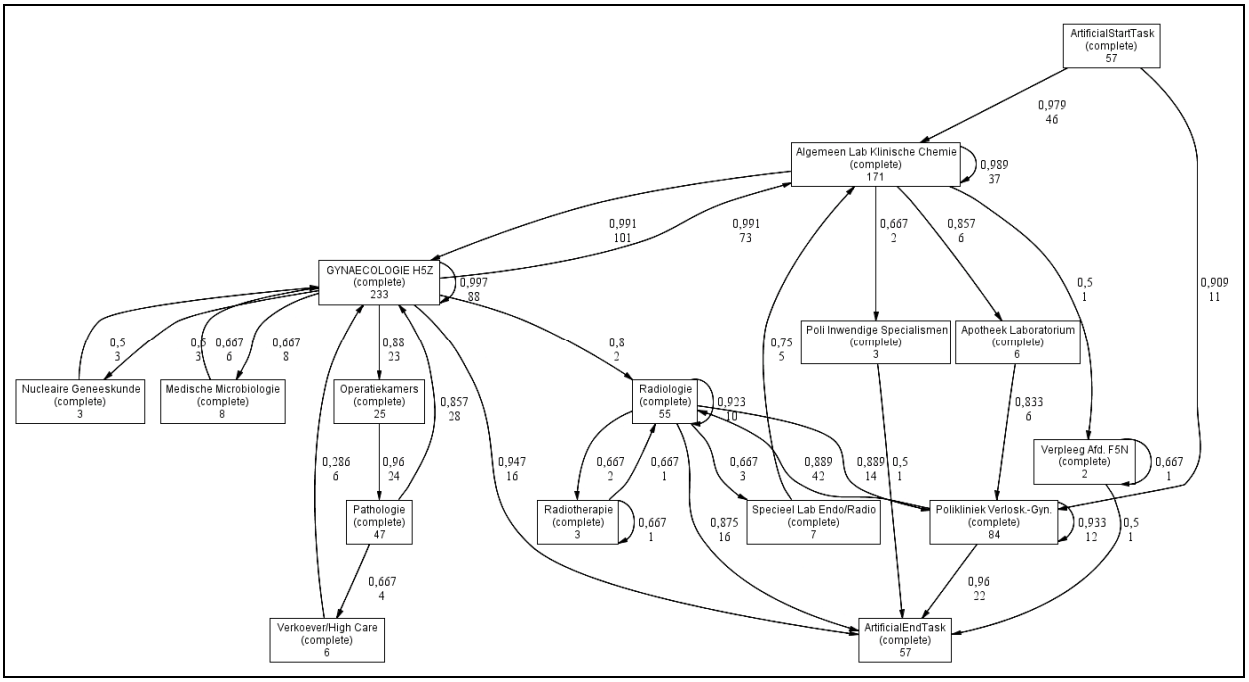


Figure 52: Heuristic miner model for the group of 57 patients in M16

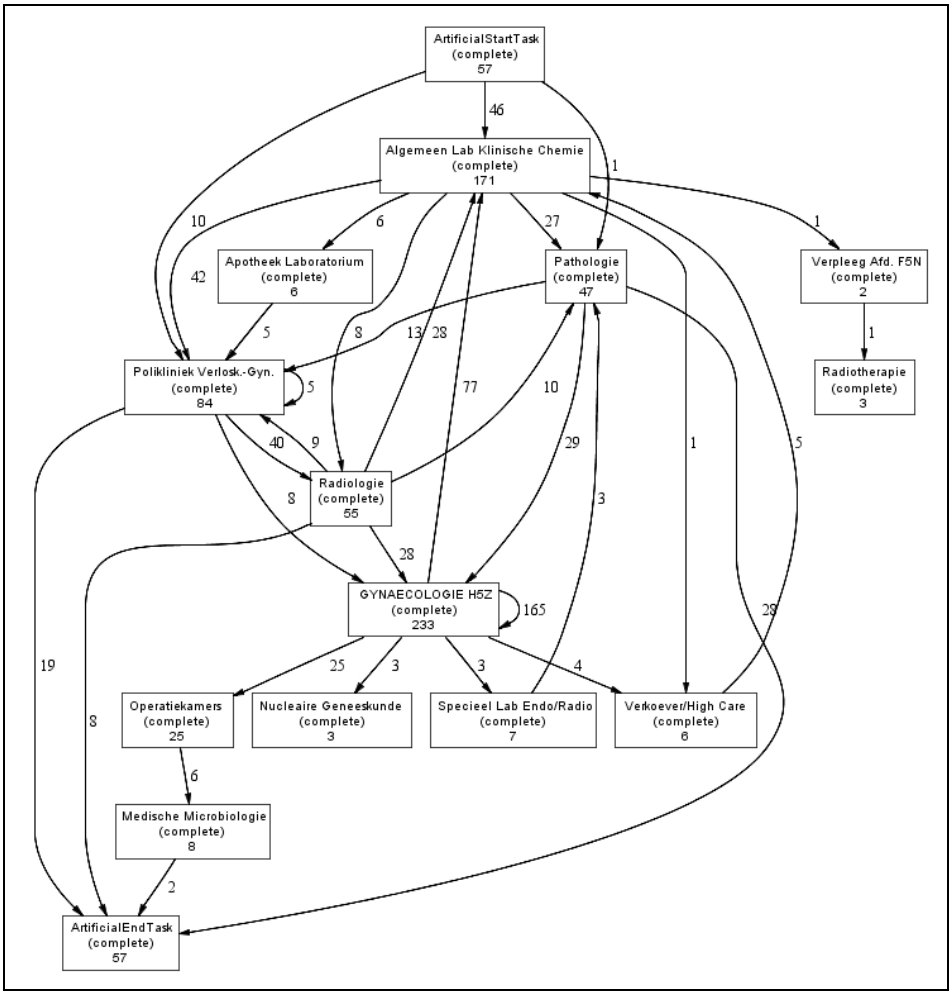


Figure 53: Genetic miner model for the group of 57 patients in M16

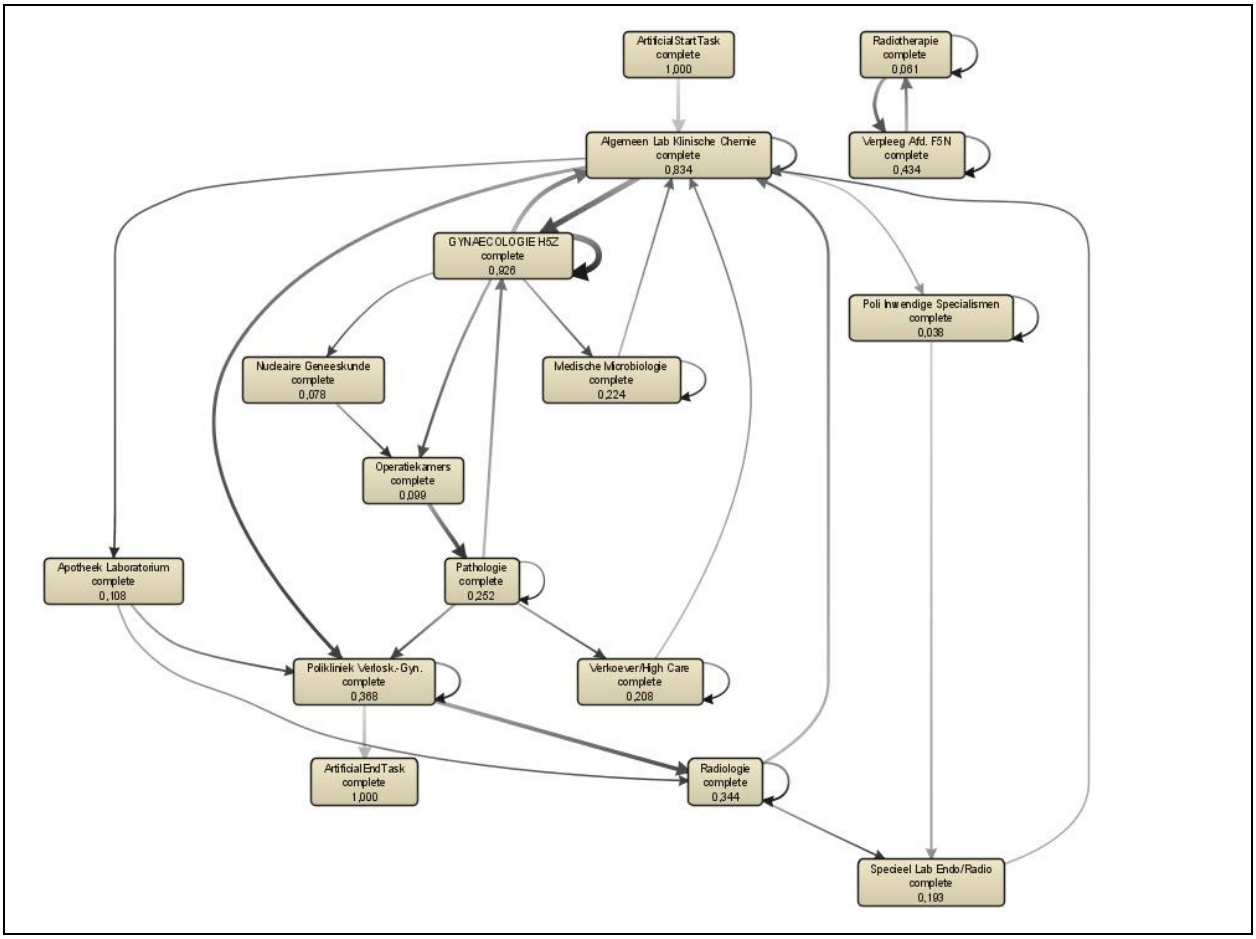


Figure 54: Fuzzy miner model for the group of 57 patients in M16

22. APPENDIX N

Based on the user interface and functionality dimensions explained in Section 5.2, it is possible to choose the best option from the analyses performed in the ProM and MagnaView tools. Furthermore, each of the analysis performed with the ProM and MagnaView tools can be categorized individually according to all the dimensions previously established in order to evaluate the benefits of each specific analysis. Next paragraphs elaborate on the detailed reasons for choosing one tool in each of the analyses cases developed in both tools as explained in Chapter 5.

- All, first, last and most frequent activities analysis results can be also obtained using the ProM tool. However, no visualization is possible to compare the results based, for example, in the DBC-code of the patients like the one obtained with MagnaView and presented in the top-left corner in Figure 55. However, the analysis with the ProM tool is possible if the medical specialists do not value the comparison of different groups of patients in a visual form. During this validation project, these comparisons were very useful in getting process information. From the evaluation of the functionality dimension included in the criteria explained in Section 5.2, the option offered by the MagnaView tool could offer more value in a healthcare environment.
- Pattern (sequential and grouped). Again, the ProM tool cannot compare results of different groups of patients in the same visualization like the ones shown in the top-right corner and bottom of Figure 55. Thus, if the comparison of results is not an important issue demanded from the analysis, the ProM tool can be used in this type of analysis. One extra issue is that when using the ProM tool, the pre-processing activities to get both visualizations (patterns sequential and patterns grouped) require more time than using the MagnaView tool. Again, based in the experience of this project, the MagnaView tool offers better understanding for these analyses.
- Patterns in the dataset. As has been shown, the developed view to find patterns in the dataset using the MagnaView tool (Figure 11) was not easily understood by the medical specialist. On the other hand, the combination of the *Dotted chart analysis* and *Performance sequence diagram analysis* plug-ins to find patterns in the dataset can be a possible way of finding useful results. These plug-ins offer interactivity with the data, extra performance indicators produced at the same time of the analysis, clear visualization, short pre-processing and time to get the results, and direct selection of events in the visualization if more focus analyses are intended. Therefore, in finding patterns in the dataset the combination of these plug-ins to find patterns in the data is a better option than using the developed view from the MagnaView tool. Future research should investigate the development of more useful visualizations in finding patterns using the visual analytics tool.
- Handover of work. This analysis is now recommended with the ProM tool due to the fact that it offers a better understanding of the dependencies between departments because it considers the causality of events. Future research could focus in the development of such characteristic using the MagnaView tool so the capacities in the comparison of events of this tool can be used to analyze the handover of work.
- Performance indicators. The performance indicators produced using the ProM tool require more pre-processing activities and offer less interaction with the visualization than the MagnaView tool. However, I believe that comparing the histograms produced with the visual analytics tool

and the process models in the ProM tool, the process mining results are a better option (only when there is no repetition of the analyzed activities on the same patient, as explained in Section 5.1.7) because they produce, automatically, more performance indicators and they offer the possibility to improve the notion of processes in a healthcare environment.

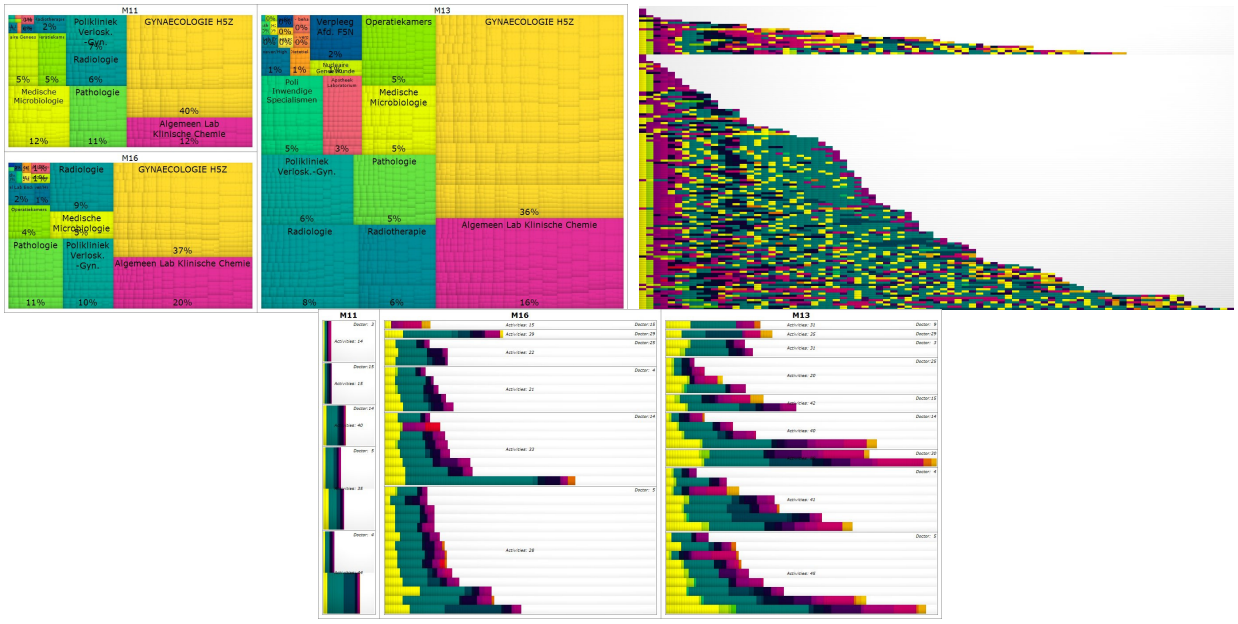


Figure 55: Examples of possible comparisons using the MagnaView tool for the all activities and patterns (sequential and grouped) views.

23. APPENDIX O

The following list contains all the activities for the different phases in the resulting method as explained in Chapter 6. These activities were produced after the analysis done with the results of the validation process in Chapter 4 and the extra analysis with the ProM tool explained in Chapter 5.

1. Build database
    1.1. Data usage
        1.1.1. Contact data manager
        1.1.2. Determine usage of data
    1.2. Extract data
        1.2.1. Identify data warehouse
        1.2.2. Data collection
        1.2.3. Integration and consolidation of data
    1.3. Transform data
        1.3.1. Pre-analysis and pre-processing of data
        1.3.2. Create database and import data
        1.3.3. Convert to MVN and MXML formats
    1.4. Load into both tools

2. Introduction session
    2.1. Before the meeting
        2.1.1. Elaborate presentation of tools and method
        2.1.2. Schedule introduction session
    2.2. During the meeting
        2.2.1. Meet specialists
        2.2.2. Determine users
        2.2.3. Determine status of change
        2.2.4. Communicate / determine goals
        2.2.5. Determine first set of KPIs and extra

3. Obtain domain knowledge

4. Preliminary analysis
    4.1. Preparation of the data
        4.1.1. LTL checker / separate DBC-codes
        4.1.2. SOM plug-in to each DBC-code
        4.1.3. Maps in MagnaView according to the identified clusters
        4.1.4. Pre-processing activities for detailed analysis
    4.2. Analysis of activities
        4.2.1. All activities
        4.2.2. First activities
        4.2.3. Last activities
    4.3. Patterns

4.3.1. Patterns (sequential)
4.3.2. Patterns (grouped)
4.3.3. Activities per day per patient
4.3.4. Patterns in dataset
4.4. Centered activity
4.4.1. Sequential
4.4.2. Grouped
4.4.3. Causal relations (1 step)
4.4.4. Causal relations (3 steps)
4.5. Process mining models
4.5.1. Heuristic miner / Genetic miner / Fuzzy miner
4.5.2. Fuzzy animation
4.6. Handover of work
4.6.1. Social network miner plug-in
4.6.2. Analyze social network plug-in
4.7. KPIs
4.8. Results
4.8.1. Elaborate presentation with the most important findings
4.8.2. Prepare questions to stakeholders about doubts in the data
4.8.3. Schedule the preliminary meeting

5. Preliminary meeting
5.1. Introduction
5.1.1. Explain method
5.1.2. Discuss initial results
5.1.3. Let actor(s) interactively "play" with tools
5.2. Selection
5.2.1. Clarify doubts about activities to include/exclude
5.2.2. Determine outliers
5.3. Determine follow up steps

6. 2nd analysis (Same activities as the preliminary analysis phase)

7. Final meeting
7.1. Before meeting
7.1.1. Elaborate presentation with the most important findings
7.1.2. Schedule the final meeting
7.2. During meeting
7.2.1. Present & discuss results
7.2.2. Receive feedback on results
7.2.3. Determine which goals were reached
7.2.4. Determine follow-up steps

8. Final documentation
    8.1. Project document
    8.2. Project files (MVN and MXML files)