Unraveling and improving the interorganizational arthrosis care process at Maastricht UMC+: an illustration of an innovative, combined application of data and process mining

K.F.Canjels[1,2], M.S.V. Imkamp[1], T.A.E.J. Boymans[1], R.J.B Vanwersch[1,2]

[1]Maastricht University Medical Center+, Maastricht, The Netherlands [2]Eindhoven University of Technology, Eindhoven, The Netherlands

rob.vanwersch@mumc.nl

Abstract. Given the forecasted major grow of osteoarthritis patients and scarcity of resources, the Maastricht UMC+ is looking for opportunities to improve the efficiency of the interorganizational care process for knee osteoarthritis patients. Currently, non-complex and complex knee osteoarthritis patients make use of the same costly facilities and highly specialized staff. By unraveling these trajectories and making use of resource substitution (especially for non-complex trajectories) substantial efficiency gains can be expected. In this report, we illustrate how an innovative data-driven three-step methodology can be used to unravel and improve the interorganizational knee osteoarthritis care process. The developed and applied three-step methodology gives guidelines on how to pre-process and integrate multiple data sets and outlines data clustering and reduction techniques that can be applied prior to process mining. We illustrate how this advanced approach supported in unraveling an initial spaghetti-like model of the complete process into easy-to-interpret sub-process models of the knee osteoarthritis care process. Moreover, we show how the subsequent analysis of these visualizations, led us to pinpoint and quantify concrete options for improving the efficiency of the knee osteoarthritis care process.

Keywords: Business process innovation, process mining, data mining

1 Introduction

Osteoarthritis is one of the largest causes of disability among elderly. Patients experience pain, instability, and limitation of movement (Doherty, Abhishek, Hunter, & Ramirez Curtis, 2017). Because of the chronic character of osteoarthritis, care crosses the boundaries of primary care (i.e. general practitioners), secondary care (e.g. general hospitals), and tertiary care (e.g. university medical centers). Since the population of patients with osteoarthritis is expected to grow with 92% over the coming 25 years in the Netherlands (Rijksinstituut Volksgezondheid en Milieu, 2018), an efficient organization of the complete interorganizational care process is of utmost importance.

This initiative focuses on improving the interorganizational care process for patients visiting the Maastricht University Medical Center+ (MUMC+) with knee osteoarthritis. Currently, non-complex knee osteoarthritis patients (e.g. patients following a short trajectory with only standardized, routine activities such as X-rays, consultations and injections) and complex knee osteoarthritis patients (e.g. patients following a long

trajectory during which many different diagnostic and treatment options have to be considered and have to be executed) make use of the same costly facilities and highly specialized staff. By unravelling these trajectories and making use of resource substitution (especially for non-complex trajectories) substantial efficiency gains can be expected. In order to identify and quantify this improvement potential, it is important to gain in-depth insights into the existing non-complex and complex trajectories. Which patient trajectories can be considered non-complex and are suitable to be seen outside the university medical center by making use of less costly facilities and staff and which are not? What is the size of these patient trajectories in terms of patient numbers? By investigating these questions, the improvement potential can be operationalized and quantified in such a way that it enables concrete implementations steps to be taken with complete support of the involved staff.

An accurate view of the current processes from beginning to end can be obtained with process mining. This is a method to discover process models based on data from event logs (e.g. billing data). Process mining has been proven useful in healthcare to optimize, among others, an emergency department (Mans, Schonenberg, Song, & van der Aalst, 2011) and stroke care (Mans, Schonenberg, Leonardi, & Panzarasa, 2008). Although process mining has proven valuable, it also faces some difficulties to produce practical models when analyzing complex processes. Patient flows are known to have a heterogenous character because patients follow different trajectories within the hospital, which leads to spaghetti models (Song, Günther, & van der Aalst, 2009). Typically, this problem is even further exaggerated in the context of interorganizational care processes, such as the knee osteoarthritis care process. Spaghetti models are hard to interpret and do not enable unraveling complex and non-complex trajectories. Hence, other analytics options beyond process mining have to be considered in the context of the knee osteoarthritis care process.

Particularly, there should be searched for ways to divide the patients in homogenous subgroups. This division reduces data complexity before process mining, enabling unraveling and improving the care process. In previous studies on health care processes clustering techniques haven been used prior to process mining (Mans, Schonenberg, Song, van der Aalst, & Bakker, 2009; Song, Yang, Siadat, & Pechenizkiy, 2013; Song et al., 2009). However, to the best of our knowledge, we are the first to report on a comprehensive technique consisting of data preparation, clustering, and data reduction techniques prior to process mining. This all applied to a process spanning the boundaries of a single organization.

In this report, we illustrate how an innovative three-step methodology can be used to unravel and improve an interorganizational care process. This methodology gives guidelines on how to pre-process and integrate multiple data sets and outlines data clustering and reduction techniques that can be applied prior to process mining. We show how the applying this innovative methodology led to the identification of substantial potential for improving the efficiency of the knee osteoarthritis care process.

The remainder of this paper is structured as follows. Chapter 2 describes the situation faced. Chapter 3 explains the approach and actions taken, followed by a discussion of the results in Chapter 4. Finally, Chapter 5 describes the lessons learned.

2   Situation faced

Given the large expected grow of osteoarthritis patients and scarcity of resources, the MUMC+ is looking for opportunities to improve the efficiency of the interorganizational care of knee osteoarthritis. The knee osteoarthritis process is complex in nature since the patients follow multiple care paths. A patient's individual care path starts with an appointment at the general practitioner (GP). When the GP decides to refer the patient to an orthopedic medical specialist, he / she can refer the patient to the outpatient city clinic or the hospital. At the city clinic, orthopedic medical specialists see patients for a consultation outside the hospital. This facilitation enables a GP to refer patients about whom he / she has doubts regarding diagnosis, treatment and / or the need to refer to hospital care. The orthopedic medical specialist at the city clinic decides on the required non-surgical or surgical treatment plan. Another option is that the GP directly refers the patient to the hospital, where a final diagnosis is set and a non-surgical or surgical treatment can be started.

In the current situation, knee osteoarthritis patients undergoing a non-complex trajectory (e.g. a short path including only standardized, routine activities such as X-rays, consultations and injections) and the ones undergoing a complex trajectory (e.g. a long path during which many different diagnostic and treatment options have to be considered and have to be executed) may either start their trajectory at the city clinic or the hospital. It is expected that even patients who start their trajectory directly at the hospital actually quite often follow a non-complex trajectory with a small amount of standardized, routine activities. Especially, the care provided to this category of patients as well as the care provided to non-complex patients currently starting their trajectory at the city clinic, might be organized more efficiently by introducing resource substitution (i.e. by making use of dedicated but less costly staff in the city clinic). In order to identify and quantify the efficiency gain potential, it is important to unravel the existing trajectories of patients. Visualizing these trajectories will provide insights into the complexity of different patient trajectories. Moreover, it allows us to analyse which facilities and staff are involved in each of the common trajectories and to evaluate whether this matches the complexity of care. Subsequently, opportunities for non-complex trajectories to be organized outside the university medical center by making use of less costly facilities and staff can be identified and quantified.

To analyze the different patient trajectories, data was collected from both the MUMC+ and the outpatient city clinic. Regarding the MUMC+, patient data was extracted from the hospital information system. For the city clinic, a list was obtained which describes the executed consultations for each patient. The patient data from the hospital information system consisted of all the activities recorded for billing purposes. This patient data consisted of 2.600 patients treated for knee osteoarthritis between January 2016 and October 2018. This data extraction resulted in 634.972 recorded events and 1.262 unique event classes (activities). The patient data from the city clinic consisted of all activities executed at this clinic. In this data set, there were only three unique event classes (activities). The two data sets were integrated and formed the input for the analysis.

Our first analysis of the process with the use of process mining resulted in a spaghetti model (Fig. **1.**). This process model is difficult to read and does not allow for unraveling complex and non-complex patient trajectories. Hence, other analytics options beyond process mining have to be considered in the context of the knee osteoarthritis care process. Particularly, we developed a three-step methodology that includes guidelines on how to pre-process and integrate multiple data sets and outlines data clustering and reduction techniques that can be applied prior to process mining.
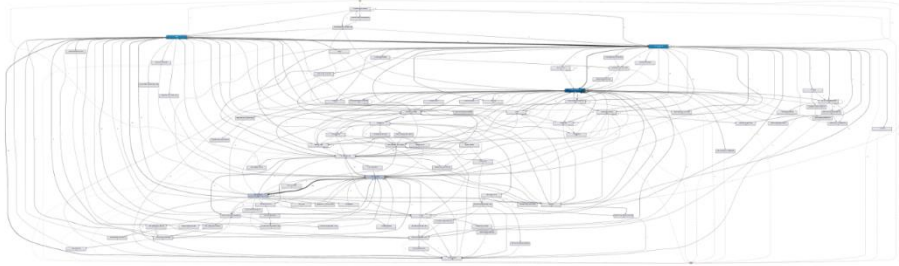

Fig. 1. The complete process model

3　　Action taken

The developed, innovative data-driven three-step methodology to enable unraveling and improving the knee osteoarthritis care process is shown in Figure 2.
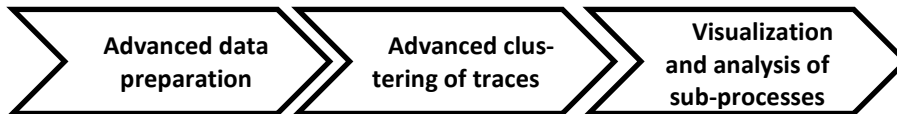

Fig. 1. A three-step methodology to analyze complex healthcare processes

As part of the first step, we discuss how to preprocess data and integrate multiple datasets, so that data and process mining techniques can be used later on to identify and visualize subgroups of patient trajectories. During the second step, we outline the application and selection of data mining techniques used to cluster traces of the complete event log. These data mining techniques enable visualization and analysis of homogenous sub-processes by means of process mining during the third step. During this initiative, all steps were executed by a data scientist and all decisions made and outcomes were discussed with an expert team consisting of three orthopaedic specialists and a process analytics expert.

3.1 Advanced data preparation

The first step "Advanced data preparation" contains five sub-steps: (1) filter relevant diagnosis, (2) merge data, (3) exclude irrelevant activities, (4) cluster activities, and (5) exclude patients with incomplete processes.

Filter relevant diagnosis: Care organizations collect data of all care activities of individual patients. When the care organization is interested in the process for a specific disease, only one or a couple of diagnoses need to be analyzed. In this case, we selected *knee osteoarthritis* and/or *loosening of the knee prothesis*, because these diagnoses are relevant in the context of knee osteoarthritis patients.

Merge data: With the focus on cross-organizational processes, files from different organizations need to be analyzed. In order to analyze the complete process, different files have to be merged. Hereby, it is important to find an identifier to link the patients from the different databases. In this case, we selected the patient number as identifier to combine the hospital and city clinic data.

Exclude irrelevant activities: In order to prevent generating a spaghetti-like model, irrelevant activities were excluded through discussion with experts. An example of an irrelevant activity in our situation is the *telcode*, which is a code required for billing purpose, indicating a consultation by phone. However, since this activity is also separately registered as *consultation by phone*, we excluded *telcode*.

Cluster activities: Some activities are interesting on a more abstract level. For example, for many diseases, the specific lab tests that are executed are not relevant. When one only needs to know whether one or more lab tests are performed at a certain laboratory, the detailed activities can be clustered and mentioned once in the model. Additionally, when multiple activities take place together, the main activities can be selected and others excluded from the model. For example, at the day of the surgery multiple activities take place (such as anesthesia, and surgery preparation). Only the main activity can be selected and the other activities can be excluded from the model while still maintaining the activity *surgery*. In addition, we indicated whether patients got physiotherapy during their stay in the hospital. For patients who received physiotherapy during their stay, we adapted the activity *admission to hospital* into *admission to hospital with physiotherapy*. For these patients, the activities corresponding to physiotherapy were excluded from the analysis, because this was now indicated by the renamed activity.

Exclude patients with incomplete processes: In order to obtain valid insights into the patient trajectories, only finished patient trajectories should be selected for inclusion. However, the selection of only finished trajectories leads to a bias towards short trajectories in recent years. This due to the fact that trajectories that have been started only recently are not likely to be completed by the end of our data analysis time horizon (October 2018). When interpreting quantitative distributions among different trajectories, one needs to be aware of this bias towards short trajectories (especially for recent time horizons). Being aware of this bias, we analyzed the percentage of patients who underwent knee surgery (long trajectories) finished their treatment process within 10 months. We found that, on average over 2016 and 2017, 86% of the patients finished their stay at the hospital within 10 months after knee surgery. Therefore we can conclude that the large majority of the patients with a long trajectory still finish their treatment within 10 months. Hence, in order to include only finished patient trajectories without a substantial bias towards short trajectories we only considered patients who started their care process before January 2018.

After executing the data preparation activities above, the resulting data should be in the form of an event log, which reflects all time-ordered activities for each patient.
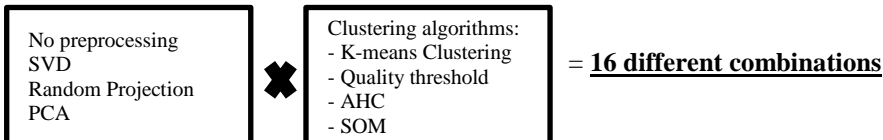
3.2 Advanced clustering of traces

During the second step, advanced trace clustering was applied to divide the data set in multiple groups with a high similarity of care activities for all patients within a group.

This was done to unravel the different patient trajectories / sub-processes and increase the understandability of the sub-process models. Clustering was performed with the trace clustering plug-in of the Process Mining Framework (ProM) version 5.2.

Clustering algorithms

Clustering algorithms are a form of unsupervised learning and cluster the data in multiple groups of similar patients to obtain partial, better understandable process models. The four trace clustering algorithms applied in this case are: K-means, Qualitative Threshold Clustering (QTC), Agglomerative Hierarchical Clustering (AHC), and Self-Organizing Maps (SOM) as described in Song, Günther. & van der Aalst (2009). As the distance measure required for calculating the dissimilarity between cases, we made use of the often-used Euclidean distance measure (Song, Günther, & van der Aalst, 2009) for all these algorithms. The four algorithms mentioned above are well known in the data mining area and have been widely applied in various domains. However, their application in the context of process mining (i.e. discovering process models) has been limited. Therefore, we compared the different algorithms to identify the most useful one in this case.

In this regard, we also looked whether the clustering algorithms might profit from dimensionality reduction techniques. The dimensionality of the data is the number of unique events which describe every record in the data (Song, Yang, Siadat, & Pechenizkiy, 2013). Trace clustering can become computationally expensive when the data dimension is high. With the use of dimensionality reduction processing time might be reduced and clustering results might be positively influenced by reducing "noise" in the dataset. In this case, we applied three well known dimensionality reduction technique: Singular Value Decomposition (SVD), Random Projection (RP), and Principal Component Analysis (PCA). This leads to a total applied number of 16 different combinations for trace clustering.

| No preprocessing SVD Random Projection PCA | ✖ | Clustering algorithms: - K-means Clustering - Quality threshold - AHC - SOM | = **<u>16 different combinations</u>** |

In addition to trace clustering, the sequence clustering algorithm is applied on the data set. Sequential clustering performs clustering based on sequential behavior of traces (Veiga & Ferreira, 2010). So, in contrast to the other techniques, sequence clustering takes explicitly the sequence of activities in the event log into account while clustering. This inclusion leads to a total of 17 combinations used to cluster the data.

Performance measure of cluster algorithms

To compare the clustering algorithms, different performance measures were used. The performance measures used are average fitness, complexity of the model, variance within clusters, and processing time. The average fitness and complexity are specifically designed for measuring the performance of the subprocess models. The average fitness describes the gap between the behavior actually observed in the log and the behavior described by the subprocess models. The complexity of the model is indicated by the size of the model in terms of nodes, arcs and the relations between them. The total variance within the cluster is calculated to indicate whether the clustering

algorithm is able to reduce the variance within the clusters compared to the total variance in the data. The performance measure processing time measures the total time required to cluster the data, and is focused on measuring the efficiency of the clustering algorithm. The performance of all algorithms was compared on all the performance measures mentioned above in order to identify the best performing ones.

Subsequently, the process mining results (in the form of visual sub-process models) of the best performing algorithms were presented to the expert team consisting of three orthopaedic specialists and a process analytics expert in order to select the clustering algorithm that led to the most easy-to-interpret grouping of patient trajectories.

3.3 Visualizations and analysis of sub-processes

After selecting the best performing clustering algorithm, the final sub-process models were visualized and analyzed with the use of process mining tool Disco during the third step with the expert team. We identified and quantified the patient trajectories that can be considered non-complex and are suitable to be seen outside the university medical center by making use of less costly facilities and staff.

4 Results

4.1 Results advanced data preparation

The advanced data preparation steps outlined in the previous section (e.g. selecting the relevant diagnosis and clustering the surgery activities) reduced the number of event classes from 1.262 to 90 in consultation with the orthopaedic specialists.

4.2 Results advanced clustering

For the advanced clustering of traces, the AHC, AHC with PCA and K-means clearly outperform the other algorithms on the different performance measures. However, when comparing the three best performing clustering algorithms (as shown in Table 1), we observe that each algorithm has its strengths and weaknesses. A darker colour in the table indicates a better performance for this measure.

Table 1. Comparison three best clustering algorithms

Performance measures	Agglomerative Hierarchical clustering	Agglomerative Hierarchical clustering with Principal Component analysis	K-means
Average Fitness	0,584	0,594	0,557
Average |A|	131,53	113,58	119,1
Average |N|	46,2	38,11	41,79
Average CNC	2,01	2,14	2,03
Average CN	86,33	76,47	78,31
Average Δ	0,091	0,102	0,095
Within Cluster Variance	10.737,51	13.169,74	10.803,73
Processing time	00:04:55	00:22:30	00:33:35

Note: |A| = number of arcs (number of transitions between activities), |N| = number of nodes (number of activities), $CNC = \frac{|A|}{|N|}$ (connectivity coefficient), $CN = |A| - |N| + 1$ (cyclomatic number), $\Delta = \frac{|A|}{|N|*(|N|-1)}$ (density), Processing time in hrs:min:sec. Note that we forced all algorithms to generate five clusters for comparison reasons. A higher number of clusters did not result in more clearly distinguishable clusters.

In order to select the final clustering algorithm, we compared the resulting subprocess models for each of the three algorithms and discussed the results with the experts. The discussion revealed that K-means performs best in coming up with easy-to-interpret groups of complex and non-complex patient trajectories. Especially, the algorithm turns out to be better in unraveling the non-surgical and surgical trajectories.

4.3 Results visualization

The chosen K-means algorithm outputs five clusters. Cluster 1 - *consultation at the city clinic* - consists of the patients who get only one consultation at the outpatient city clinic. Cluster 2 - *consultation at MUMC+* - consists of the patients who only get one consultation in the hospital, where after their trajectory is finished. Cluster 3 – *consultation at MUMC+ with X-ray* – consists of patients who get an X-ray and consultation in the hospital. Cluster 4 – *conservative treatment at MUMC+* - consists of the patients with a conservative treatment path in the hospital, the patients in this cluster get on average 3.94 activities. Finally, cluster 5 - *surgical treatment at MUMC+* - is the most complex group and consists of patients who get surgery in the hospital. These patients have the most extensive treatment path with on average 24.32 activities.

As an example, the clusters *consultation at MUMC+ with X-ray* (cluster 3) and *conservative treatment at MUMC+* (cluster 4) are visualized and explained. The cluster *consultation at MUMC+ with X-ray*, as shown in Figure 3, represents patients with a non-complex care path. The Figure shows that all patients enter the hospital with an X-ray of the knee (X-Knie in Dutch), after which they get a first consultation (1e consult algemeen). After this first consultation, they will leave the hospital and finish their trajectory. The cluster *conservative treatment at MUMC+*, as shown in Figure 4, represents the patients with a longer conservative treatment path. It is shown that most patients start with an X-ray of the knee (X-Knie) and end with a follow-up consultation (Vervolgconsult algemeen). The X-ray of the knee is typically followed by the first consultation (1e consult algemeen). After this consultation, there are several options. Many patients continue with a follow-up consultation (Vervolgconsult algemeen, 255), but they might also get an injection (injectie kenacortinspuiting, 182) or a knee MRI (mri knie, 103). After the first follow-up consultation, there is a split between the patients, 281 patients leave the hospital, but other patients come back for multiple follow-up consultations or injections before finishing their trajectory.
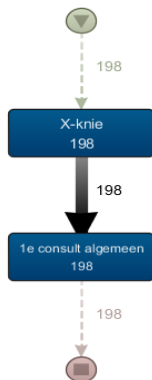

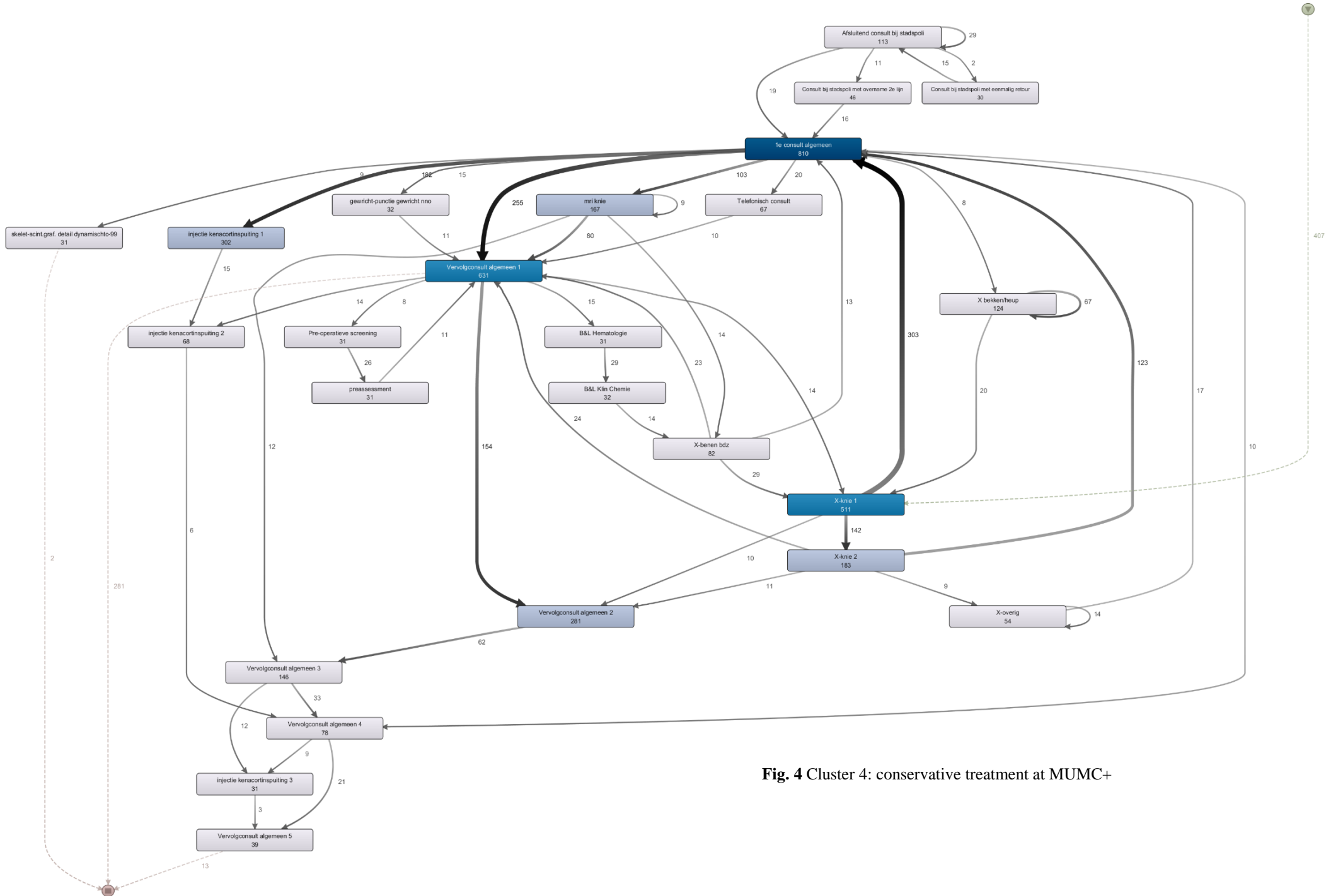**Fig. 3.** Cluster 3: consultation at MUMC+ with X-ray.

Fig. 4 Cluster 4: conservative treatment at MUMC+

4.4 Improvement opportunities

Analyzing the subprocess models of the fives clusters, reveals that three clusters (*Consultation at the city clinic, Consultation at MUMC+,* and *Consultation MUMC+ with X-ray*) can be classified as non-complex patient trajectories, because they represent patients who got at most one X-ray and one consultation at the outpatient city clinic or hospital. All these trajectories are suitable to be completely executed in the city clinic where patients can be seen by dedicated but less expensive staff (i.e. physician assistants, specialized GPs or physiotherapists under supervision of medical specialists). The cluster *conservative treatment at MUMC+* (cluster 4) is slightly more complex because it consists of the patients with a longer conservative treatment path in the hospital. However, through the visualization of the process model it is found that the majority of the group (583 of the 891 patients) only has standardized, routine activities (consultations, X-rays and injections). Again, these trajectories can be completely executed in the city clinic where patients can be seen by dedicated but less expensive staff. Also, patients within cluster 4 that do not follow a complete standardized, routine trajectory could start the first part of their trajectory (i.e. X-ray and consultation) in this new city clinic context. The most complex care is received by patients in the cluster *surgical treatment at MUMC+* (cluster 5). Despite the complexity of this group visual inspection leads us to conclude that patients belonging to cluster 5 can start their trajectory (i.e. X-ray and consultation) in the new city clinic context. The effects of the redesign options discussed above are discussed in Table 2.

Table 2. Impact redesign

	Current facility			Potential facility	
	Full trajectory in city clinic	**Start in city clinic, follow-up in MUMC+**	**Full trajectory in MUMC+**	**Full trajectory in city clinic 2.0***	**Start in city clinic, follow-up in MUMC+**
Cluster 1 - *Consultation at the city clinic*	151 (9.4%)	0	0	151 (9.4%)	0
Cluster 2 - *Consultation at MUMC+*	0	0	163 (10.1%)	163 (10.1%)	0
Cluster 3 – *Consultation at MUMC+ with X-ray*	0	0	198 (12.3%)	198 (12.3%)	0
Cluster 4 - *Conservative treatment at MUMC+*	43 (2.7%)	55 (3.4%)	793 (49.1%)	583 (36.1%)	308 (19.1%)
Cluster 5 - *Surgical treatment at MUMC+*	0	9 (0.6%)	201 (13%)	0	210 (13%)
Total	194 (12.0%)	64 (4.0%)	1355 (84.0%)	1095 (67.9%)	518 (32.1%)

Note: In the city clinic 2.0 resource substitution is introduced (patients will be seen by dedicated but less expensive staff as compared to the current situation in the city clinic).

As shown in Table 2, the percentage of patients that undergo their full (non-complex) care trajectory in the city clinic is expected to grow from 12.0% to 67.9%. Also, the percentage of patients that undergo their initial non-complex start of their trajectory in the city clinic is expected to grow from 4.0% to 32.1%. In the new situation, there is potential for introducing resource substitution in the city clinic, i.e. patients will be seen by dedicated but less expensive staff (i.e. physician assistants, specialized GPs or physiotherapists under supervision of medical specialists). Given the annual forecasted patient numbers, the above percentages and the average number of activities per cluster that can be moved out of the hospital context, approximately 1.250 patient sessions (consultations and/or injections) can be moved out of the hospital and can be organized more efficiently in the setting of the city clinic, leading to an expected yearly efficiency gain of at least €75.000. Moreover, patients are likely to benefit from a reduction of average waiting times by unraveling non-complex (fast queue) and complex (slow queue) trajectories and from being (more often) seen in the more patient-friendly setting of a city clinic.

5   Lessons learned

Healthcare organizations are increasingly facing pressure to improve the efficiency of their care processes. As such, they are challenged to provide adequate care for patients at lower costs. In this regard, it is important that the different levels of care-complexity of patients match with involved facilities and resources. By means of applying a comprehensive data and process mining approach, we were able to unravel non-complex and complex trajectories of knee osteoarthrosis patients and pinpoint opportunities for improvement. In our study, five clusters of patient trajectories were identified: three clusters of non-complex care (*Consultation at the city clinic, Consultation at MUMC+,* and *Consultation at MUMC+ with X-ray*), one with a majority of non-complex care (*Conservative treatment at MUMC+*), and one of complex care (*Surgical treatment at MUMC+*).

In correspondence with our expectations, the resulting models showed the possibility to facilitate care for a large group of patients outside the university medical center where they can be seen by less expensive resources and staff. We identified that the percentage of patients that are able to undergo their full or partial (non-complex) care trajectory in an efficient city clinic context is expected to grow substantially. Beside the substantial efficiency gain that can be realized, a reduction of average waiting times (especially for non-complex patients) is expected as well as an increase in patient satisfaction. The results revealed that the identification and quantification of complex and non-complex patient groups is an important asset for the improvement of the healthcare process as it supports the "willingness for change" of the involved staff.

> *"Now, we finally see the size of the non-complex patient group and the possible impact of reorganizing our care processes"*
> Orthopedic surgeon, MUMC+

From a methodological perspective, this study clearly indicates the value of the developed data-driven three-step methodology for unraveling and improving (care) processes. The results show the potential of thorough pre-processing of data and making

use of data mining tools prior to process mining. The application showed that a spaghetti-like mined model can be transformed into easy-to-interpret sub-process models by making use of appropriate pre-processing of data and data mining techniques. Other parties might benefit from the applied methodology to analyze and improve similar cross-organizational healthcare processes. Also, this methodology could be extended to complex processes outside the healthcare setting. In order to foster further uptake, it is recommended to focus future research on developing guidelines for selecting the best performing clustering algorithm. Given the current absence of these guidelines in literature, we had to perform a quite time-consuming evaluation of multiple cluster algorithms in order to select the most suitable one. Despite this limitation, this case illustrates clearly the process innovation potential of the thorough, combined application of data and process mining.

References

Doherty, M., Abhishek, A., Hunter, D., & Ramirez Curtis, M. (2017). Clinical manifestations and diagnosis of osteoarthritis. *UpToDate*.

Mans, R., Schonenberg, H., Leonardi, G., Panzarasa, S., Cavallini, A., Quaglini, S., Aalst, W. van der. (2008). Process mining techniques: an application to stroke care. *Studies in health technology and informatics*,*136*, 573–578.

Mans, R. S., Schonenberg, H., Song, M., Aalst, W. M. P. van der, & Bakker, P. J. M. (2009). Application of process mining in healthcare: a case study in a Dutch hospital. *Communications in Computer and Information Science*, *25,* 425-438.

Song, M., Yang, H., Siadat, S. H., & Pechenizkiy, M. (2013). A comparative study of dimensionality reduction techniques to enhance trace clustering performances. *Expert Systems with Applications*,*40*(9), 3722–3737.

Song M., Günther C.W., Aalst W.M.P. van der. (2009). Trace clustering in process mining. *Lecture Notes in Business Information Processing, 17*, 109-120.

Veiga, G.M., & Ferreira, D.R. (2010). Understanding spaghetti models with sequence clustering for ProM. *Lecture Notes in Business Information Processing*, *43*, 92–103.