Towards cost-effective treatment of periprosthetic joint infection: from statistical analysis to Markov models

Yulia E. Kaliberda[1], Vasiliy N. Leonenko[1[0000-0001-7070-6584]] and Vasiliy A. Artyukh[2]

[1] ITMO University, Kronverksky Pr. 49A, St. Petersburg, 197101, Russia
[2] Russian Scientific Research Institute of Traumatology and Orthopedics named after R.R. Vreden, Akademika Baykova st, 8, 195427, St. Petersburg, Russia
vnleonenko@yandex.ru

Abstract. The aim of the research is to perform statistical analysis and to build probabilistic models for the treatment of periprosthetic joint infection (PJI) based on available data. We assessed and compared the effectiveness of different treatment procedures from the terms of the objective result (successful PJI treatment without relapse) and the subjective assessment of the condition of the patients (Harris Hip Score). The ways to create prognostic models and analyze cost-effectiveness of treatment strategies are discussed based on the results obtained.

Keywords: Periprosthetic Joint Infection, Statistical Analysis, Markov Models, Python.

1 Introduction

Periprosthetic joint infection (PJI) is a serious complication which may occur after arthroplasty [1]. It is associated with high morbidity and requires complex treatment strategies including multiple surgical revisions and long-term antimicrobial treatment, because the implant as a foreign body increases the pathogenicity of bacteria and the presence of biofilm makes the diagnosis and treatment problematic [2]. The investigations related to the creation of cost-effective approaches to PJI treatment [3,4] mention the issues connected with the corresponding analysis due to lack of quality studies.

Beside the problem of finding the best PJI treatment method in general, relying on both direct (percentage of successful PJI elimination) and indirect treatment outcomes (such as resulting increase in quality of life of the patients who underwent the treatment), there is an arising challenge of finding an optimal treatment strategy in advance for the particular patient based on his individual characteristics [5]. This challenge became actual as a part of the personalized medicine concept and requires the research based on multidisciplinary approach and relying on statistical analysis, mathematical modeling and machine learning [6]. The ultimate aim of that direction of research consists in developing a computational tool to predict the consequences of a fixed treatment strategy for a given patient. Such a tool, when put into use by the

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

healthcare professionals, will help enhance the quality of PJI treatment both in terms of cost-effectiveness and the quality of life of individuals undergoing treatment.

The aim of the research described in the presented paper was to perform statistical analysis of the PJI treatment methods effectiveness used in clinical practice of Russian Scientific Research Institute of Traumatology and Orthopedics named after R.R. Vreden (St Petersburg, Russia) and to formulate a discrete-event stochastic model for the clinical paths of the patients exposed to PJI, taking into account both the treatment procedure itself and its long-term treatment consequences, including the relapse of PJI and the related interventions. We assessed and compared the effectiveness of different treatment procedures from the terms of the objective result (successful PJI treatment without relapse) and the subjective assessment of the condition of the patients (Harris Hip Score). The ways to create prognostic models and analyze cost-effectiveness of treatment strategies are discussed based on the results obtained.

2 Data

The dataset used in the analysis contains the records of 609 patients who received treatment for PJI infection in the period of 2000-2020. The patient records are divided into two observation groups – the retrospective and the prospective one. The retrospective group was composed of the disease histories taken from the archives, whereas the prospective group was filled by constantly adding the data of the patients currently undergoing treatment starting from 2014. The structure of the dataset with respect to treatment methods is shown in Fig.1.

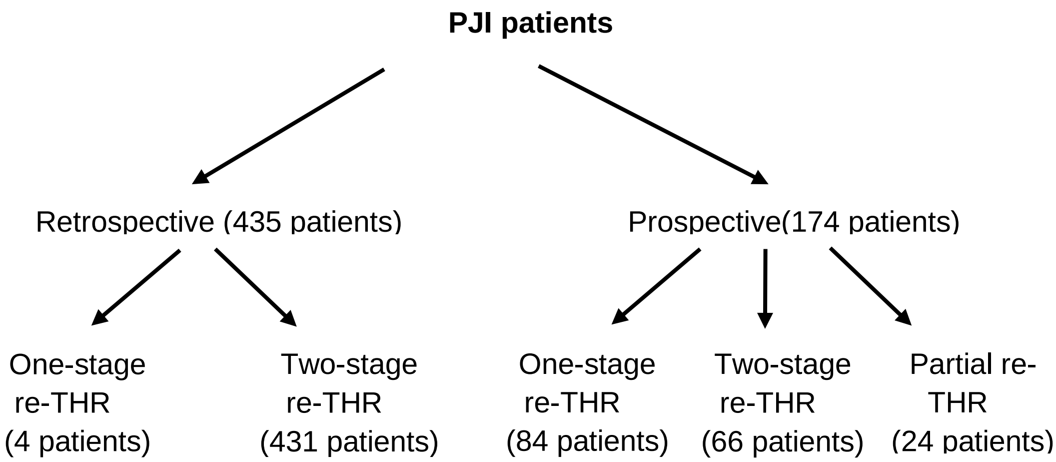

Fig. 1 – The dataset structure related to the treatment methods. Re-THR stands for the reoperation of total hip replacement

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

In retrospective group, the applied treatment methods included one-stage re-endoprosthesis and two-stage re-endoprosthesis, also known as reoperations of total hip replacement (re-THR). Since the set with one-stage re-THR patients contained only 4 records, we excluded it from further observation. In the prospective group, a new method was presented in addition to the two named before, which was partial re-THR. The two consecutive interventions of two-stage re-THR in the retrospective group were separated by more than 2 months, while in the prospective group waiting time between surgeries did not exceed 2 months. The two-stage re-THR prospective group was divided into 2-3 weeks waiting time and 6-8 weeks waiting time. All people were being observed till the end of 2020.

In patients' data for each patient there are defined features:

- Name, birthdate
- Operation dates (the total number of operations varied between 1 and 10)
- Operation types
- Two sets of Harris Hip Scores (HHS) [7], - before and after treatment, - used to measure the quality of life of the patient
- The resulting state of the patient related to PJI, measured during his last attendance to healthcare services (PJI relapse or no PJI)

In the ideal situation, the prospected number of operations performed on each patient is fixed and defined solely by the PJI treatment method, but in many cases additional operations were required due to the relapse of PJI or other issues (postoperative wound hematomas, spacer dislocations, etc). There exists 15 different types of operations in overall, which could be divided into three groups, depending on how the particular operation type was connected with PJI:

- Operations which have no connection with PJI
  - Endoprosthesis (EP) installation + spacer removal
  - EP installation (no spacer)
  - Non-infectious: spacer dislocation
  - Other (suturing, etc.)
  - Non-infectious: periprosthetic fracture case
  - Unknown
- First case of PJI or PJI relapse
  - Debridement (DAIR)
  - Debridement + spacer installation
  - EP components replacement + debridement
  - Debridement + full EP replacement
  - Joint drainage + long-term suppressive antibiotic therapy (ABT)
- PJI relapse
  - Debridement + spacer reinstallation
  - Disarticulation
  - Spacer removal + support osteotomy
  - Debridement + support osteotomy + Girdlestone resection arthroplasty
  - Joint drainage

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

Further on, we do not distinguish the types of operations and consider them equal (see more comments on the matter in Section 4).

3 Statistical analysis

3.1 Comparing the treatment outcomes

The first aim of the statistical analysis was to compare the effectiveness of different treatment methods in terms of two indicators available from the data:

- The ratio of PJI relapse cases among the patients (reflects the objective result of the treatment method application)
- The improvement of Harris Hip Score according to the answers of the patients (reflects the increase in the quality of life as a result of performing the treatment procedures).

The corresponding proportions and confidence intervals were calculated using the algorithm implemented in Python 3.6. The results are presented in Table 1 and Fig. 2.

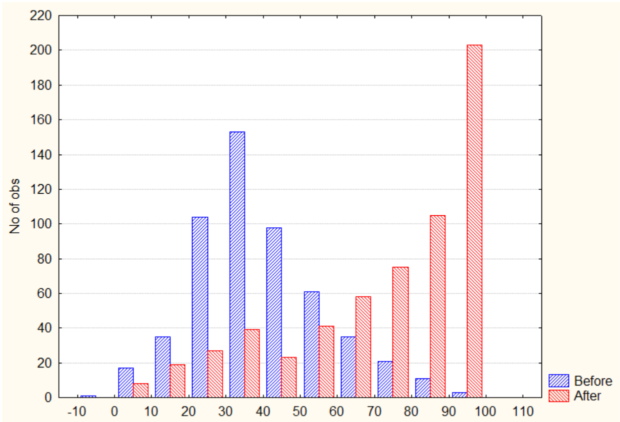

Fig. 2 – The frequency of particular Harris Hip Scores among the patients before (blue) and after the treatment (red).

Table 1 - Harris score mean confidence intervals and PJI relapse ratio for each treatment type. First questioning was made when a patient was admitted for a treatment, second questioning corresponds to current patients' score.

Group name	**HHS before treatment**	**HHS after treatment**	**Proportion of people with PJI**
Two-stage	(37.3685, 40.8680)	(66.4771, 71.6734)	0.092

> 2 months			
One-stage	(33.2760, 40.8489)	(73.6527, 84.4723)	0.110
Two-stage 2-3 weeks	(32.3068, 64.4205)	(56.6777, 104.0496)	0.077
Two-stage 6-8 weeks	(39.5684, 50.8899)	(79.4123, 89.7127)	0.060
Partial	(37.5556, 55.5277)	(60.6142, 83.3858)	0.091

Fig. 2 demonstrates that the distribution of Harris scores after the treatment compared to the one before confirms the increase of functional capacity, since it is shifted to the right by its median and is left-skewed. The difference was proved to be statistically significant by the Sign test ($p<0.05$), applied to the whole sample and to the subsamples corresponding to each of the five treatment methods. As it can be seen from Table 1, the lowest PJI proportion (0.06) is shown by a two-stage 6-8 weeks treatment method. This method may also boast the second highest median Harris score (see Fig. 3). The best median HHS and the second best proportion of PJI-free patients corresponds to two-stage 2-3 weeks treatment method, however more data might needed to verify this result since the sample of patients for the method is not big enough.

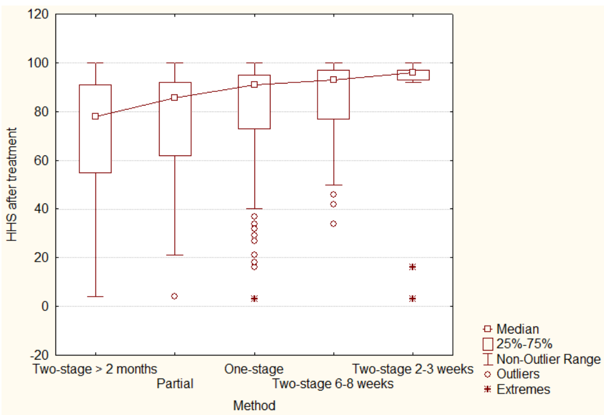

Fig. 3 – Distribution of Harris scores after the treatment

3.2 Quantifying the role of multiple operations on the well-being of patients

As it was mentioned earlier, in the course of treatment additional operations might be required, which are caused by the necessity to deal with PJI relapse and other un-

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

expected situations. Since every additional surgical intervention might negatively influence both the PJI treatment outcome and the functional capacity of the patients, it is important to analyze the dynamics of the corresponding indicators. Thus we counted the dependence of PJI ratio on the number of operations for different treatment methods - Table 2 shows the example for the retrospective group (Two-stage re-THR, > 2 months).

Table 2 – Operation connections with PJI stage in retrospective group

Number of operations in the group	Number of patients in the group	Number of patients without PJI	PJI relapse proportion
1	211	202	0.04265
2	121	109	0.09917
3	58	48	0.17241
4	25	21	0.16000
5	13	10	0.23077
6	5	3	0.40000
7	2	2	0.00000
8	2	2	0.00000
9	0	0	N/D
10	3	1	0.66667

Such transformation serves to a future target of an overall work - creating a probabilistic model for each group of patients.

The dependence between the number of operations and mean Harris score is shown in Table 3, while Figure 4 demonstrates the distribution of Harris scores. The average statistic was calculated in both cases without division on treatment methods due to the small sample sizes.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

Table 3 – Average Harris scores depending on the number of operations

Number of operations in the group	**Average HHS after the treatment**	**Average HHS improvement**	**Number of patients in the group**
1	74.432	33.382	273
2	75.383	36.605	187
3	69.375	27.344	75
4	60.851	29.000	30
5	43.800	9.100	14
6	45.500	9.750	5
7	3.000	-18.000	3
8	N/D	N/D	2
9	N/D	N/D	0
10	39.666	-16.334	3

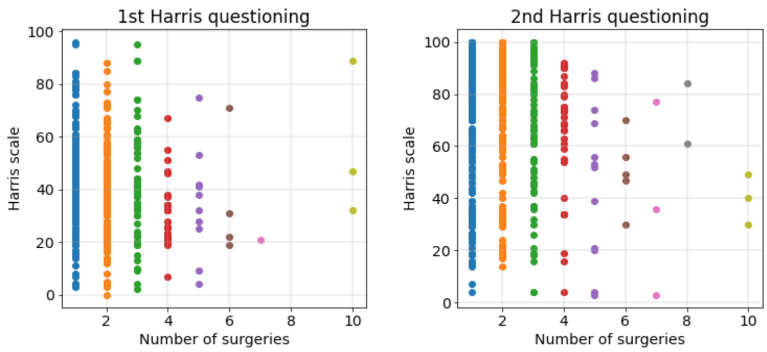

Fig. 4. Harris scale results before and after treatment by patient groups distinguished by overall resulting number of operations

DOI: 10.1007/978-3-030-77967-2_41

It is seen from the Tables 2 and 3 that PJI ratio is an increasing function of the resulting operation number while Harris score tends to decrease. In other words, according to the data, the more operations a particular patient had, the bigger risk of PJI he has and the lower functional capacity he retains.

It is important to note that the number of patients who has 5 operations and more is rather small (see Tables 2 and 3), which might alter the correctness of the conclusions. Another issue that might potentially decrease the quality of comparison of the methods (particularly, the comparison of final PJI ratios) and the subsequent model calibration on data was the difference in the time periods when different treatment methods were first introduced. As a result, the older methods might have records with longer observation time compared to the newer ones. If a patient in the record is listed as one without PJI relapse, having had few operations, it might mean that he was indeed effectively treated by the method, or, alternatively, that the sufficient time after the treatment has not passed to register the subsequent relapse and start treating it with the help of surgical manipulations. As a result of analysis of observation times in each treatment group (Fig. 5), we have concluded that although their distributions are indeed different, the mean observation times do not differ dramatically (with a two-stage >2 months group being somewhat an exception).

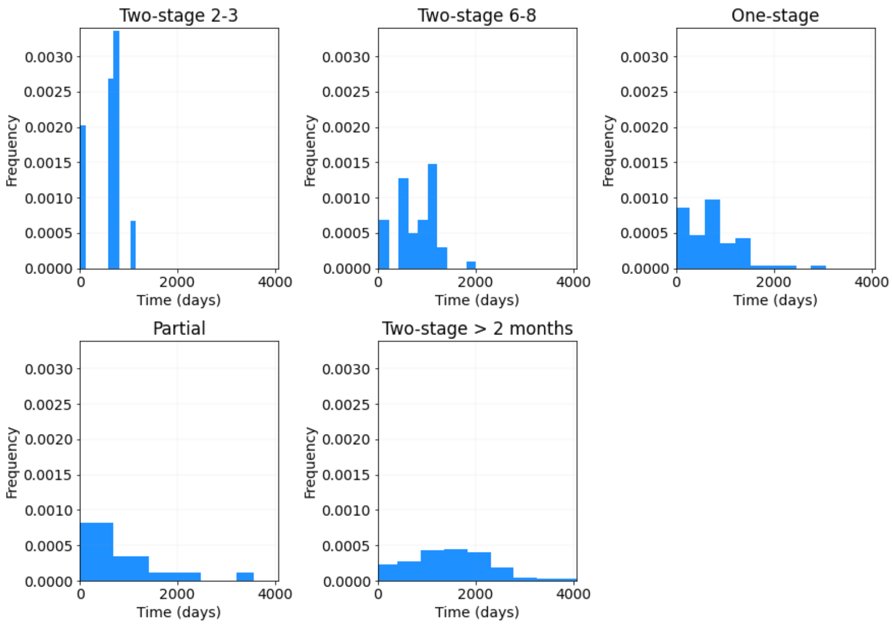

Fig. 5. Observation time distribution plots

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

4 Model

After performing the statistical analysis of the treatment outcomes, the next step of the study was to formulate a Markov model which could be used in predicting the individual patient trajectories in case of PJI treatment. There is a number of the corresponding models published [8-11] which differ by their structure. The additional data analysis was performed to assess the applicability of different concepts to our situation. The following model features were analyzed.

Generalized states [8,9] vs explicit states [10]. In some of the proposed Markov models there are no explicit states taking into the account consecutive numerous operations after the first surgical intervention. For instance, in [8] and other similar studies all the operations after the first unsuccessful revision are generalized in one "Re-revision" state which might lead to death or recovery. In our case, since we have found a dependence of a number of operations on the patients' well-being and we wanted to have a model capable of predicting those indicators, we decided to explicitly state all the operations one by one along with the probabilities that after operation N an operation N+1 might be required. At the same time, we have not distinguished the operation types as it was made in [10]. Particularly, we have not used the information on whether the operation was PJI-related or not (see Section 2).

The selected approach has several advantages and disadvantages:

- Since the majority of considered operations imply aggressive surgery, we assumed that they equally might affect patients' condition – this was supported by the analysis of Harris scores and PJI ratios in the previous section.
- To expand the study with the cost-effectiveness assessment, which is planned in the future, it is necessary to take into account all the operations performed – so aggregating the states would cause issues with the cost calculation.
- On the other hand, there exist a mandatory number of surgical interventions connected to the treatment methods, which is not taken into account and might affect the method comparison results and the accuracy of models (for instance, two-stage re-THR has a planned additional operation by default compared to one stage re-THR, with the transition probability between the stages close to 1).

Time-dependent [11] vs time-independent state transitions. The transition probabilities between the model states might be calculated in two ways:

- The probability of transition per a fixed time period (essentially a rate)
- The probability of transition *per sé* without regarding time

In the former case the model explicitly includes time, whereas in the latter the time "jumps" according with the state transitions, so those jumps are not equal in length. This might affect the accuracy of treatment method comparison in case when the average time period between the states is substantially different for different methods. For instance, the method which in average requires PJI relapse treatment not earlier

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

than in five years after the first PJI treatment should be considered more effective compared with the method which causes a PJI relapse after five months. At the same time, it is more beneficial to perform all the necessary operations within the minimal time interval, because it enhances the quality of life of a patient. In our case we decided to disregard these nuances, because the distribution of time between the operations (Table 4, Figure 6) seem to be quite similar among the prospective methods. In case of the retrospective method (two-stage > 2 months) it is twice as big, which additionally discourages from using this method.

Table 4 – Confidence intervals in days for the average time between the operation for each treatment type

Group name	Confidence interval	Mean	Size
Two-stage 2-3 weeks re-THR	(402.613, 802.617)	602.615	13
Two-stage 6-8 weeks re-THR	(658.482, 895.910)	777.196	81
One-stage re-THR	(583.224, 826.704)	704.964	53
Two-stage > 2 months	(1402.159, 1567.156)	1484.657	435
Partial re-THR	(476.210, 1186.707)	831.458	24

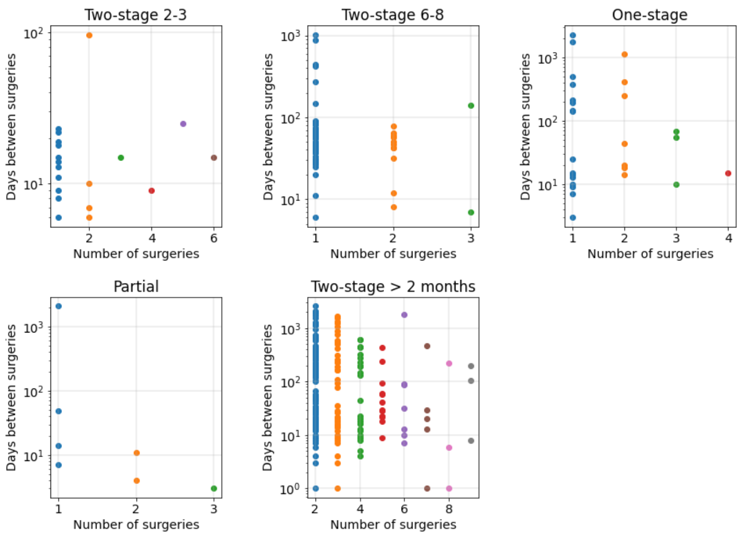

Fig. 6. Plots for times between surgeries for each treatment group

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

As a result of the mentioned considerations, we have chosen to develop a Markov model which complies to the following principles:

- The possible model states are “Operation 1”, “Operation 2”,…, “Operation N”, “PJI” and “No PJI”.
- Each patient proceeds through the sequence of operations until he dies or his monitoring stops for other reasons
- The state “PJI”/ “No PJI” is final and reflects the condition of the patient before his death or at the moment of final monitoring event

Probabilistic trees which correspond to modeling the transitions of patients between the Markov model states for three different treatment methods are demonstrated in Figures 7-8. These trees are used to predict the trajectory of a fixed patient undergoing a certain treatment, with the help of Monte Carlo methods. Model implementation is performed via Python programming language. We have selected the trees for “one-stage re-THR” and “two-stage re-THR with 6-8 weeks” for the demonstration purposes, because the corresponding samples are big enough to hope for the correct estimation of the transition probabilities (apart from “Two-stage 2-3 weeks re-THR” and “Partial re-THR”) and also they are in active use these days (apart from “Two-stage > 2 months”). The model structure is not dependent on the treatment method type and can be further verified when the new data will become available.

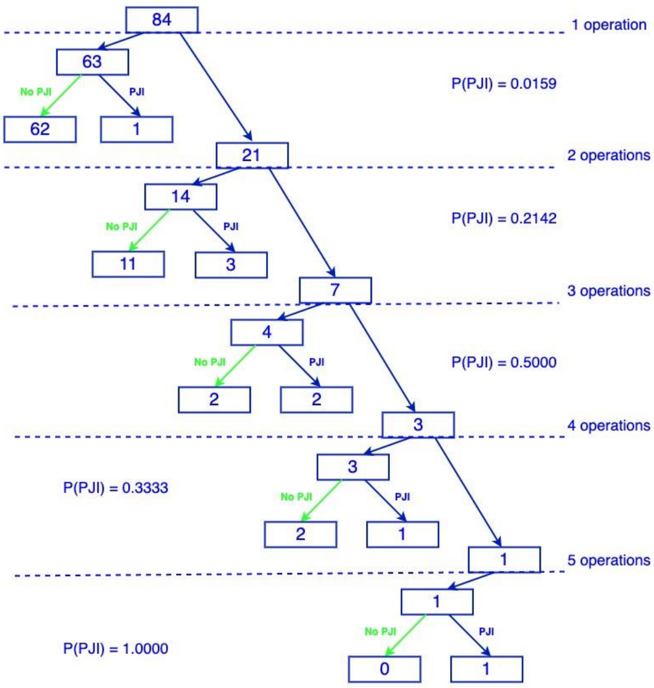

Fig. 7 – Results of the treatment with one-stage re-THR

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

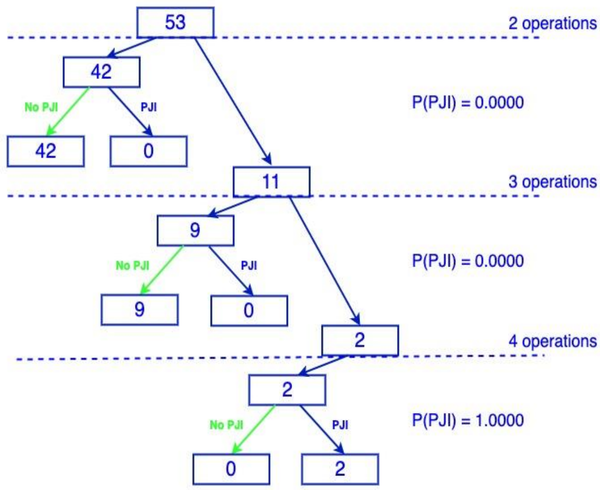

Fig. 8 - Results of the treatment with two-stage re-THR with 6-8 weeks between operations

5 Discussion

In the current study, we have analyzed the data on PJI treatment and proposed a Markov model to reflect the transition between the treatment stages for the particular patient, with transition probabilities depending on the treatment method. The size of the sample under study somewhat limits the possibility to make explicit conclusions from the analysis as to comparative effectiveness of the treatment methods. It was however shown that the treatment itself enhances the functional capacity of the patients, and that this capacity is badly affected by repetitive operations, not depending on treatment method. Also the PJI relapse chance correlates positively with the number of operations performed (although this result should be considered *cum grano salis* due to necessity of providing a meaningful interpretation to it).

The Markov model presented in the study serves as a first step towards the prediction of patient trajectories. The absence of a timer in our model might affect the perspectives of cost-effectiveness analysis (in case of evaluating characteristics such as QALY gained per year). Also, calculating time between the model states might be necessary for the application of the model in the decision support system, so in the future the model structure might be reconsidered.

The issue worth noticing is that the transitions between the operations have a complex nature and are not always directly connected with treatment method effectiveness. For instance, repetitive operations due to PJI relapse are the example of the process connected with the disease course, and more operations mean less efficient treatment. At the same time, some operations follow each other in a regular fashion (like two stages in PJI treatment), they are planned according to the schedule and

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

there should not be any probabilistic transitions between the corresponding model states. Finally, some of the operations are caused by accidents, which are random but not connected to PJI (for instance, spacer dislocation). They should be considered an external factor. The model which will distinguish these three groups of situations will be more correct in terms of describing the patient treatment dynamics.

In addition to the mentioned model improvement, future directions of the study will include cost-effectiveness analysis based on the cost of particular operations, and alongside it the consideration of the model with explicit operation states. The obtained study results related to changes in the quality of life of patients might be enhanced with the usage of EQ-5D scale beside Harris score. Finally, the classification of clinical pathways is to be done using CLIPIX software [12].

6 Acknowledgement

This research is financially supported by The Russian Science Foundation, Agreement #19-11-00326.

References

1. Assmann, G., Kasch, R., Maher, C. G., Hofer, A., Barz, T., Merk, H., & Flessa, S. (2014). Comparison of health care costs between aseptic and two stage septic hip revision. *The Journal of arthroplasty*, *29*(10), 1925-1931.
2. Masters, E. A., Trombetta, R. P., de Mesy Bentley, K. L., Boyce, B. F., Gill, A. L., Gill, S. R., ... & Muthukrishnan, G. (2019). Evolving concepts in bone infection: redefining "biofilm","acute vs. chronic osteomyelitis","the immune proteome" and "local antibiotic therapy". *Bone research*, *7*(1), 1-18.
3. Li C, Renz N, Trampuz A. "Management of Periprosthetic Joint Infection" // Hip Pelvis. 138-146 (2018)
4. Hernández-Vaquero D, Fernández-Fairen M, Torres A, Menzie AM, Fernández-Carreira JM, Murcia-Mazon A, Guerado E, Merzthal L. "Treatment of periprosthetic infections: an economic analysis" // ScientificWorldJournal. 2013 May 28;2013:821650.
5. Kovalchuk, S. V., Funkner, A. A., Metsker, O. G., & Yakovlev, A. N. (2018). Simulation of patient flow in multiple healthcare units using process and data mining techniques for model identification. *Journal of biomedical informatics*, *82*, 128-142.
6. Kovalchuk, S. V., Knyazkov, K. V., Syomov, I. I., Yakovlev, A. N., & Boukhanovsky, A. V. (2015). Personalized clinical decision support with complex hospital-level modelling. *Procedia Computer Science*, *66*, 392-401.
7. Harris WH. Traumatic arthritis of the hip after dislocation and acetabular fractures: treatment by mold arthroplasty. An end-result study using a new method of result evaluation. J Bone Joint Surg Am 1969; 51:737–55.
8. Tack, P., Victor, J., Gemmel, P., & Annemans, L. (2021). Do custom 3D-printed revision acetabular implants provide enough value to justify the additional costs? The health-economic comparison of a new porous 3D-printed hip implant for revision arthroplasty of Paprosky type 3B acetabular defects and its closest alternative. Orthopaedics & Traumatology: Surgery & Research, 107(1), 102600.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41

9. Parisi, T. J., Konopka, J. F., & Bedair, H. S. (2017). What is the long-term economic societal effect of periprosthetic infections after THA? A Markov analysis. *Clinical Orthopaedics and Related Research®*, *475*(7), 1891-1900.
10. Wolf, C. F., Gu, N. Y., Doctor, J. N., Manner, P. A., & Leopold, S. S. (2011). Comparison of one and two-stage revision of total hip arthroplasty complicated by infection: a Markov expected-utility decision analysis. *JBJS*, *93*(7), 631-639.
11. Fisman, D. N., Reilly, D. T., Karchmer, A. W., & Goldie, S. J. (2001). Clinical effectiveness and cost-effectiveness of 2 management strategies for infected total hip arthroplasty in the elderly. *Clinical Infectious Diseases*, *32*(3), 419-430.
12. Funkner, A. A., Yakovlev, A. N., & Kovalchuk, S. V. (2017). Towards evolutionary discovery of typical clinical pathways in electronic health records. *Procedia computer science*, *119*, 234-244.

ICCS Camera Ready Version 2021
To cite this paper please use the final published version:
DOI: 10.1007/978-3-030-77967-2_41