Association for Information Systems

AIS Electronic Library (AISeL)

Research-in-Progress Papers

ECIS 2017 Proceedings

Spring 6-10-2017

PREDICTING THE DURATION OF SURGERIES TO IMPROVE PROCESS EFFICIENCY IN HOSPITALS

Martin Riekert
University of Hohenheim, Germany, martin.riekert@uni-hohenheim.de

Marc Premm
University of Hohenheim, Germany, marc.premm@uni-hohenheim.de

Achim Klein
University of Hohenheim, Germany, achim.klein@uni-hohenheim.de

Lyubomir Kirilov
University of Hohenheim, Germany, lyubomir.kirilov@uni-hohenheim.de

Hannes Kenngott
University of Heidelberg, Germany,, hannes.kenngott@med.uni-heidelberg.de

See next page for additional authors

Recommended Citation

Riekert, Martin; Premm, Marc; Klein, Achim; Kirilov, Lyubomir; Kenngott, Hannes; Apitz, Martin; Wagner, Martin; and Ternes, Lena, (2017). "PREDICTING THE DURATION OF SURGERIES TO IMPROVE PROCESS EFFICIENCY IN HOSPITALS". In Proceedings of the 25th European Conference on Information Systems (ECIS), Guimarães, Portugal, June 5-10, 2017 (pp. 2842-2851). ISBN 978-0-9915567-0-0 Research-in-Progress Papers.
http://aisel.aisnet.org/ecis2017_rip/33

Authors
Martin Riekert, Marc Premm, Achim Klein, Lyubomir Kirilov, Hannes Kenngott, Martin Apitz, Martin Wagner, and Lena Ternes

PREDICTING THE DURATION OF SURGERIES TO IMPROVE PROCESS EFFICIENCY IN HOSPITALS

Research in Progress

Riekert, Martin, University of Hohenheim, Stuttgart, Germany, martin.riekert@uni-hohenheim.de

Premm, Marc, University of Hohenheim, Stuttgart, Germany, marc.premm@uni-hohenheim.de

Klein, Achim, University of Hohenheim, Stuttgart, Germany, achim.klein@uni-hohenheim.de

Kirilov, Lyubomir, University of Hohenheim, Stuttgart, Germany, lyubomir.kirilov@uni-hohenheim.de

Kenngott, Hannes, University of Heidelberg, Heidelberg, Germany, hannes.kenngott@med.uni-heidelberg.de

Apitz, Martin, University of Heidelberg, Heidelberg, Germany, martin.apitz@med.uni-heidelberg.de

Wagner, Martin, University of Heidelberg, Heidelberg, Germany, martin.wagner@med.uni-heidelberg.de

Ternes, Lena, University of Heidelberg, Heidelberg, Germany, lena.ternes@stud.uni-heidelberg.de

Abstract

Predicting the duration of surgeries is an important task because of the many dependencies between surgery processes and the hospital processes within other departments. Thus, accurate predictions allow for better coordinating patient processes throughout the hospital. Prior data-driven research provides evidence for accurate predictions of surgery durations enhancing the efficiency of surgery schedules. However, the current prediction models require large sets of features, which make their adoption more intricate. Moreover, prediction models focus on the surgery department and neglect potential effects on other departments. We use a unique dataset of about 17,000 surgeries to study how particular features and machine learning algorithms affect the prediction accuracy of major surgery steps. The prediction models that we study require few features and are easy to apply. The empirical findings can be useful for the design of surgery scheduling systems.

Keywords: Surgery Duration Prediction, Machine Learning, Business Process Management.

1 Introduction

Business process management (BPM) in hospitals involving surgeries is complex because multiple interdependent departments are affected (Armony et al., 2015). The departments are highly autonomous and often dominated by the expensive surgery department (Macario, 2010). Due to limited surgery resources, the business process greatly depends on preceding processes of other patients. Thus, process parts in other departments (such as the wards, anesthesia, recovery or intensive care) must be often suspended, are prone to workarounds (Tucker et al., 2014; Röder et al., 2015), and have to react in an ad-hoc manner to requests from the surgery department. A parallel process in surgery scenarios is the request for the next patient as soon as the end of the concurrent surgery is foreseeable (Figure 1). Requesting the next patient too late means its process of being transported to the surgery department and the anesthetization will cause the operating room (OR) to idle waiting for a patient. However, requesting the patient too early causes unnecessarily long anesthesia time and may have spillover effects on the resource and location utilization. The optimization of triggering this request usually involves knowledge about coordinating the allocation of operating rooms and resources (Schultz et al., 2007; Jebali and Diabat, 2015; Saadouli et al., 2015). An accurate prediction of process step durations may be used to adapt the triggering of patient requests accordingly and optimize timing of patient requests as well as post-surgery processes. To improve process efficiency in terms of physical and human resource utilization as well as patient well-being, this work addresses the problem of designing accurate prediction models for the major steps of a patient's process in the surgery department.

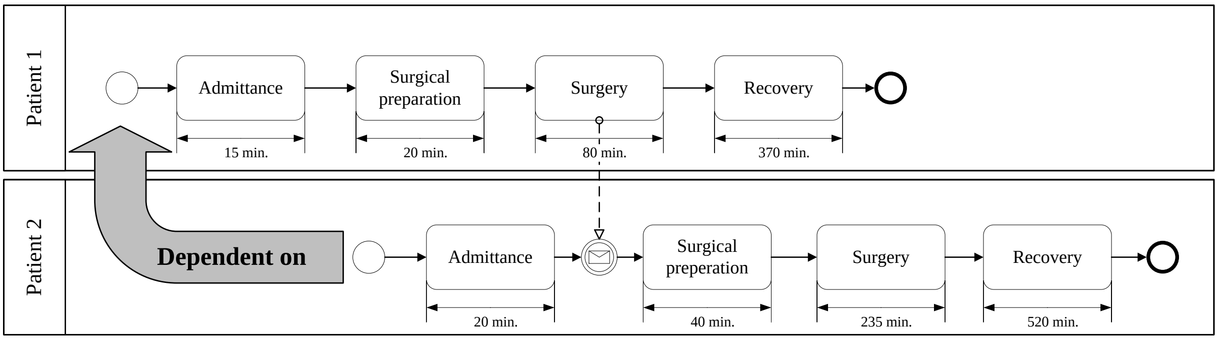

Figure 1. Dependencies between patient processes.

Previous studies focused on predicting the surgery duration using the patient's age, gender, body mass index, and associated risk, the surgeon, the anesthesiologist, the disease, and the procedure type (Cardoen et al., 2010; Eijkemans et al., 2010; Gomes et al., 2012; Lowndes et al., 2016). While some studies managed to explain about two-thirds of the variation of surgery durations (ShahabiKargar et al., 2014; Kayış et al., 2015) other's prediction models explained up to 81% of the variance (Gomes et al., 2012). However, practical utility is limited due to the large number of required features, forming the inputs of the prediction models. In practice, the inputs might not be easily available for process managers, especially for emergency surgeries that have to be taken into account immediately (Lamiri et al., 2008). In this setting, ad-hoc prediction of the process is key to effective process management and schedules should be adapted dynamically. Prior research has partly addressed this research gap (van Dongen et al., 2008; Van der Aalst et al., 2011; Polato et al., 2014; Polato et al., 2016). However, these studies do not use features that are specific to the surgery process and do not compare different machine learning (ML) algorithms. Considering prior research, we formulate the research question:

What is the effect of features and machine learning algorithms on the explained variance (R^2) of the duration of process steps?

The outline of the paper is as follows: Section 2 presents our research model. Section 3 reports on the dataset and methods used. Section 4 presents the results, which are then discussed in section 5. Finally, section 6 concludes.

2 Research Model

The research model of our study is shown in Figure 2. The machine learning-based models are designed for predicting the major steps of the operational business process in the surgery department. The outcomes of experiments are evaluated using the R^2 metric, which is defined as the percentage of correctly explained variation of the duration of a process step.

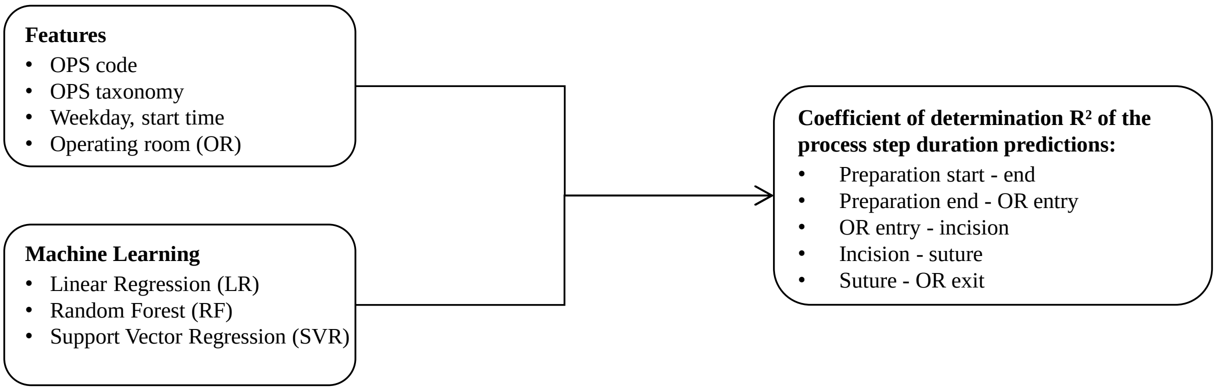

Figure 2. Research model.

OPS code: We use the official German classification for surgeries, procedures and general medical actions as an adaptation of the international classifications of the procedures in medicine (ICPM) of the World Health Organization (WHO), termed OPS - German Procedure Classification[1]. Li et al. (2009) studied the predictive power of the operating procedure, which describes the individual procedures performed during a surgery. They found that regression models, which only use operating procedures as an explanatory variable, outperform both historical averages for a given department as well as models that assume that surgery durations are log-normally distributed. Also several other studies found the operating procedure to be an important feature for surgery time prediction (Eijkemans et al., 2010; Gomes et al., 2012; Master et al., 2016). We assume that adding the operating procedure to the features increases the R^2 of the surgery duration prediction.

OPS taxonomy: Additionally, we posit that the OPS code bears further predictive power due to its taxonomic structure. The OPS code begins with the superclass at the left side of the code and becomes more specific towards the right side, e.g., “5-448.00” consists of the subclasses “5: surgical procedures”, “5-44: other operations at the stomach”, “5-448: other reconstructions at the stomach” and “5-448.42: fundoplication: laparoscopic”. This procedural classification schema can be used for feature engineering. The rationale is that there are so many different procedures that many process codes are observed only once or twice in the whole dataset. In this case, the OPS super class provides information about the business process duration in terms of the most similar operating procedures. Therefore, we assume that adding the OPS taxonomy to the OPS code increases the R^2 of all our predictions.

Weekday and preparation start time: Further factors that we posit to impact the duration of a surgery are the time and the weekday of a surgery. For example, Stepaniak et al. (2010) found the time of the day to be a relevant and statistically significant factor for predicting medical operation times alongside the experience and the age of the surgeon. However, their results are mixed with some medical procedures taking less time in the morning while others being shorter in the afternoon. Kayis et al. (2012) also explicitly tested the predictive power of temporal factors such as time of the day, day of the week, month, and year but only found significant results for surgeries performed in January as well as after 5 pm. In a follow up study Kayis et al. (2015) reiterated their results with respect to surgeries, performed after 5 p.m. Furthermore, they found that the first surgery for the day in a given operating room is shorter than if the same surgery was performed later that day as well as durations of surgeries performed in June, July and October significantly differ from durations of surgeries, performed throughout the rest of the year. Based on previous findings, we assume that adding the weekday and surgery time to the features increases the R^2 of the surgery duration prediction.

[1] https://www.dimdi.de/static/en/klassi/ops/

Operating room: It has been also suggested that the location, which is chosen for a surgery, could contain information about the expected complexity of the surgery and the required resources, e.g. procedures, performed in the Ambulatory Procedure Unit, tend to be shorter and simpler (Master et al., 2016). In addition, Kayis et al. (2015) found evidence that the operating times in particular operating rooms in a large US-based hospital significantly differ from others. Therefore, we expect that the surgery location increases the R^2 of the surgery duration prediction.

Machine Learning: Examining the relationship between machine learning algorithms and the resulting model quality for surgery durations is necessary due to the no free lunch theorem for machine learning, which states that there is no a priori superior machine learning algorithm (Wolpert, 1996). First, linear regression (LR) has been used as an algorithm in several studies (Eijkemans et al., 2010; Gomes et al., 2012; ShahabiKargar et al., 2014; Hosseini et al., 2015; Thiels et al., 2017). Second, research for surgery time prediction has shown that random forests (RF) (Breiman, 2001) outperform most other machine learning algorithms and it can increase the prediction accuracy by 30-35% compared to traditional methods used in healthcare facilities (Gomes et al., 2012; ShahabiKargar et al., 2014). Third, recent literature relies on Support Vector Regression (SVR) (Drucker et al., 1997; Smola and Scholkopf, 2004) for process time prediction (Polato et al., 2016). Therefore, we conclude that advanced machine learning methods should achieve higher R^2 than linear regression.

3 Method

The prediction models of surgery durations were based on a large dataset from a German university hospital. The dataset consisted of over 17,000 surgeries, which were recorded over three years by the hospital staff. The operations were manually recorded during and after the surgery procedure, contain emergency cases and are associated with timestamps, indicating the beginning and the end of a process step. The duration between the timestamps served as target variables for the machine learning algorithms. For the duration of each process step, Table 1 provides the descriptive statistics of each process step. The process steps are described in the following: (1) “preparation start – end" is the duration of the process step that prepares the patient for the surgery. It includes the preparation and induction of the narcosis. (2) “preparation end – OR entry”: in this process step the patient is transported to the OR. (3) “OR entry – incision”: in this process step the physician prepares the patient for surgery, sterility is set up and the Team-Time-Out is performed. (4) “incision – suture”: in this process step the surgery is conducted. (5) "Suture - OR exit": in this process step a spot in the recovery room is reserved, narcosis is stopped, the patient is relocated and the patient is transported to the recovery room.

Process step	**N**	**M**	**SD**	**Q1**	**Q2**	**Q3**
Preparation start - end	17,537	34.7	11.4	30	35	45
Preparation end - OR entry	12,288	46.4	39.9	15	35	65
OR entry - incision	17,435	27.5	15.1	18	25	34
Incision - suture	17,694	110.7	104.4	28	70	166
Suture - OR exit	17,424	18.0	14.3	8	15	25

Table 1. Descriptive statistics of each process step in minutes. N: sample size, M: mean, SD: standard deviation, Q1: first quartile, Q2: second quartile/median, Q3: third quartile.

The dataset was cleaned using the following procedures. We removed all operating timestamps that had no OPS code, had no value or were higher than the default cutoff of three standard deviations. “Preparation end – OR entry” had fewer values because emergency surgeries may have surgery preparations in the OR. This results in negative preparation end - OR entry durations, which we removed from the dataset. The features for the prediction of the surgery process duration were: the weekday, the time when the surgery preparation started (start time), the OR, the OPS code and the OPS taxonomy.

The following preprocessing was applied to the dataset. We used the OPS code, the weekday and the time the preparation of the surgery started in minutes as features for the machine learning algorithms. The start time was treated as a metric value and was mapped to one feature. For the categorical features weekday and OPS code one-hot-encoding was used. For the OPS taxonomy, we also applied feature engineering. The OPS code string was cut after three characters (e.g., “5-4”) for the first level of the taxonomy, four characters (e.g., “5-44”) for the second level of the taxonomy and five characters (e.g. “5-448”) for the third level of the taxonomy. The complete OPS code (e.g., “5-448.42”) is not included in the OPS taxonomy. We scaled the numerical features by subtracting the median and dividing by the interquartile range of the features' values.

The OPS codes are often recorded for reimbursement after the surgery was conducted, which is also the case in our dataset. We can assume that the exact surgery procedure is often not known in advance so that the planned OPS code and the actual OPS code differ. Therefore, the probable additional information of the OPS code requires further investigation towards the extent the missing information prior to the surgery affects the prediction. We examined this by comparing the prediction results using the OPS taxonomy (first to third level only) vs. using the OPS taxonomy and the complete OPS code. The justification is that, for example, a surgery that is classified at the third level of the taxonomy as “5-448: other reconstructions at the stomach” will rarely become a subset of the class “5-447 revision after gastric resection”. Thus, the third level taxonomy of the OPS code should be the same before and after the surgery. We also provide the R^2 of the model using the even more general first and second taxonomy level (e.g., “5-44: other operations at the stomach”).

We applied the machine learning algorithms with their default configurations as provided in the scikit-learn machine learning library (Pedregosa et al., 2011) and increased the number of trees to 150, following previous work (Oshiro et al., 2016). For model selection, we calculated the R^2 metric out of sample using 10-fold cross validation, which has been found to be better suited than bootstrapping (Kohavi, 1995).

4 Results

This section reports results of prediction models for surgery process step durations using different features and machine learning algorithms. Table 2 provides the results of all prediction models. The best prediction models achieved a R^2 of 0.73 for the incision – suture duration, a R^2 of 0.349 for the preparation start – end duration, a R^2 of 0.325 for the suture – OR exit duration and R^2 of 0.199 for the preparation end – OR entry duration. The explained variance of the OR entry – incision duration by the best performing model was low (R^2 = 0.069), leading us to conclude that the prediction of this surgery process step duration is not possible with the underlying dataset. In the following, we assess the impact of the features and machine learning algorithms.

OPS code: We assumed that the OPS code is an important feature for the prediction model and has a positive impact on R^2. Table 2 indicates that OPS as a single feature is the most important driver of the duration of the incision – suture process part and achieves R^2=0.71 with linear regression.

OPS taxonomy: We expected that adding the OPS taxonomy to the OPS code increases R^2. The addition of the OPS taxonomy to the OPS code {OPS-tx, OPS} increased R^2 over the OPS code {OPS} for all process steps and machine learning algorithms. For example, the R^2 for the random forest using only OPS increased from 0.692 to 0.723 after adding the {OPS-tx}. Furthermore, we investigated the deviation between the preoperative OPS code and the postoperative OPS code. The impact on the overall performance is demonstrated by comparing the prediction model of the OPS taxonomy {OPS, OPS-tx} with two scenarios where in scenario 1 only the first and second level of the taxonomy {OPS-tx level 1 and 2} and in scenario 2 the first, second and third level of the taxonomy {OPS-tx} are available before all surgeries. For both scenarios the process step incision – suture was the only one to show a R^2 difference of over 0.05. For this step, random forest resulted in a maximum R^2 difference of 0.14 with R^2 = 0.582 for {OPS-tx level 1 and 2} vs. R^2=0.723 for {OPS, OPS-tx} and a maximum R^2

difference of 0.06 with R^2=0.668 for {OPS-tx} vs. R^2=0.723 for {OPS, OPS-tx}. Both scenarios indicate that excluding the postoperative information of the full OPS code in our prediction models should neither increase the R^2 by more than 0.15 for the incision – suture duration nor reduce the R^2 for the other durations by more than 0.05.

Weekday and preparation start time: We enquire whether adding the surgery time and weekday to the features further increases the R^2 of the prediction. The R^2 of the models using the features weekday and start time {weekday, start time} are relatively low or negative in the out-of-sample evaluation. However, the prediction of the process step preparation start-end is improved from R^2=0.190 for {OPS-tx, OPS} to R^2=0.287 for {OPS, OPS-tx, weekday, start time} using random forest.

Operating room: We examine the predictive power of the operating room feature for the process step durations. Table 2 indicates the operating room {OR} to be a good predictor (R^2=0.325 for LR and RF) for the suture – OR exit duration. Furthermore, the location feature combined with all other features {OPS, OPS-tx, weekday, start time, OR} with linear regression achieves the highest R^2 of 0.727 for the incision – suture duration.

Machine learning algorithms: Here, we investigate whether RF and SVR can explain more variance than LR. RF achieved the highest results for the preparation start – end duration (R^2=0.349) and preparation end – OR entry duration (R^2=0.199). LR achieves the highest R^2 of 0.727 for incision – suture duration. RF and LR achieved equal results for the suture - OR exit duration R^2=0.325. SVR is the worst performing method of the 3 methods investigated in this study.

	Features	ML	Preparation start – end	Preparation end – OR entry	OR entry – incision	Incision – Suture	Suture – OR exit
OPS code	{OPS}	RF	0.187	0.122	0.063	0.692	0.27
		LR	0.170	0.135	0.047	0.710	0.274
		SVR	0.139	0.099	0.041	0.496	0.238
OPS taxonomy	{OPS-tx level 1 and 2}	RF	0.142	0.162	0.064	0.582	0.299
		LR	0.142	0.162	0.064	0.582	0.299
		SVR	0.049	0.111	0.009	0.555	0.247
	{OPS-tx}	RF	0.185	0.170	0.077	0.668	0.309
		LR	0.183	0.168	0.074	0.667	0.308
		SVR	0.099	0.123	0.029	0.626	0.259
	{OPS-tx, OPS}	RF	0.19	0.152	0.067	0.723	0.291
		LR	0.174	0.141	0.048	0.719	0.276
		SVR	0.154	0.137	0.052	0.673	0.27
Weekday, start time	{start time, weekday}	RF	-0.091	-0.007	-0.011	0.092	0.183
		LR	0.011	0.014	0.007	0.035	0.096
		SVR	-0.024	-0.047	-0.06	-0.107	0.086
	{OPS, OPS-tx, weekday, start time}	RF	0.287	0.154	-0.035	0.682	0.200
		LR	0.215	0.142	0.049	0.72	0.276
		SVR	0.234	0.14	0.056	0.676	0.271
Operating room	{OR}	RF	0.197	0.146	0.049	0.436	**0.325**
		LR	0.196	0.145	0.049	0.436	**0.325**
		SVR	0.147	0.099	-0.022	0.402	0.262
	{OPS, OPS-tx, OR}	RF	0.233	0.125	0.040	0.718	0.271
		LR	0.24	0.149	0.061	0.726	0.295
		SVR	0.235	0.148	0.066	0.689	0.301
All	{OPS, OPS-tx, weekday, start time, OR}	RF	**0.349**	**0.199**	-0.007	0.694	0.244
		LR	0.267	0.152	0.063	**0.727**	0.295
		SVR	0.284	0.152	**0.069**	0.691	0.304

Table 2. Descriptive statistics of the R^2 metric for predictions of the process step durations. Bold are the highest results. OPS-tx: OPS taxonomy, start time: time the surgery preparation started, weekday: day of the week the surgery took place, ML: machine learning algorithm.

5 Discussion

We find that predicting the process step durations surrounding the incision – suture duration is possible with R^2=0.349 for the preparation start – end duration, R^2=0.325 for the suture – OR exit duration and R^2=0.199 for the preparation end – OR entry duration. The best prediction model achieved a R^2 of 0.73 for the duration of the incision – suture process step. This result is comparable to previous work (Eijkemans et al., 2010; Gomes et al., 2012).

The relevant features according to our research are OPS code, OPS taxonomy and operating room, whereas the efficacy of the feature start time and weekday is mixed. First, the most decisive feature is the OPS code, which is a consistent finding to previous work (Li et al., 2009). Second, the OPS taxonomy slightly improves the duration predictions of all process steps by providing more general features, which allow the predictions of rare surgery procedures and may reduce overfitting of the machine learning algorithms (Hastie et al., 2009). Third, the operating room is a feature with few different categories, but major impact on suture – or exit durations. This finding could be explained by specific devices that are only installed in specific rooms that increase the time until the patient leaves the OR. Fourth, our findings concerning mixed results for start time and weekday contradict previous work, which used the start time as a feature (Kayış et al., 2012, 2015). A possible explanation for the small impact of weekday and start time on the duration of the process steps might be that processes in the observed hospital changed several times. Fifth, the R^2 of the prediction models vary depending on the chosen machine learning algorithms *and* the predicted process steps. The no free lunch theorem provides a possible explanation for this result, i.e. there is no universally best machine learning algorithm (Wolpert, 1996).

Contributions: First, research on the prediction of surgery durations from a BPM perspective is still scarce and previous research has only predicted incision – suture duration or the duration the patient stayed in the OR. We extend prior work by predicting process step durations, allowing surrounding stations to better assess the status of the OR due to more specific estimates for each process step. Second, we introduce the OPS taxonomy as a new feature for process step duration prediction that requires no additional data and increases the R^2 of the prediction models. The OPS taxonomy has previously been used in context of process step identification (Meier et al., 2015). Third, by focusing on few features, we sustain patient and physician privacy and nevertheless achieve R^2 comparable to previous work (Gomes et al., 2012).

Implications for IS research: First, our results indicate that predictions of almost all process step durations depend on features that are easy to record and extract and similar features are available for other business processes. Therefore, this study could be expanded by adding further process steps in the hospital or even by generalizing to other organizations. Second, research gaps exist towards more accurate predictions of the process durations. Exploring new features, applying feature selection (Guyon and Elisseeff, 2003; Chandrashekar and Sahin, 2014), applying regularization to the linear regression algorithm (Friedman et al., 2010) or training other machine learning algorithms could help to increase the models' R^2. Third, other process parameters may be predictable and relevant to BPM. Such parameters could be costs, optimal amount of personnel per process step or required supplementing resources.

Implications for practice: Our study suggests that preemptive planning of business processes can improve business process performance by aligning dependent business processes. The results of the study indicate that accurate prediction models for the incision – suture duration can support operating room coordination. Surgeries can be planned with higher reliability even if only few features are available for surgery scheduling. We also demonstrated that process steps before and after the incision – suture duration can be predicted, which could improve the process management further.

Limitations: First, in our dataset the OPS code was recorded after the surgery was conducted. Therefore, we estimated the impact by constructing features that contained less information than the least amount of preoperative information available. Our results indicate that the postoperative information has not a strong impact on R^2. However, a test set of preoperative OPS codes should be used in future work. Second, we note that the no-free lunch theorem in machine learning limits the generalizability of our results (Wolpert, 1996). This work's dataset may be different from other hospitals' data and a generalization may not be entirely possible. Third, the predictions assume that nothing in the hospital changed since the dataset was recorded, i.e., future changes of organizational structure or operating procedures are unaccounted for in the prediction models. Fourth, we did not investigate the organizational structure concerning correlations that might result from scheduling policies that would try to conduct complicated surgeries in particular rooms or schedule predictable surgeries first. Fifth, our current experiments do not provide statistical hypothesis tests. Further studies should go beyond the evaluation in this paper. Sixth, whereas the research model was largely supported, our predictions achieved only moderate R^2 except for the prediction of the duration of the incision – suture process step, which is the most important step in practice.

6 Conclusion

Our work foremost contributes to IS research by better understanding the factors affecting the duration of steps of the operational business process of patients in hospitals. Our research model contains five design factors and was tested by experimental evaluation using a unique dataset of over 17,000 surgeries. The prediction models address durations of all major steps of operational processes involving patients in the operating room. Our experimental evaluation achieves high R^2 of 0.73 for the duration for the most important incision – suture process step. However, in contrast to other models, it requires only few features. Thus, this research helps to shorten the path for practitioners and researchers to design accurate prediction models to support process managers in hospitals. These results should be considered for efficient BPM in hospitals and for reducing waiting times for surgery. The code used in our experiments can be retrieved from https://wi2.uni-hohenheim.de/analytics.

Acknowledgements

This work has been supported by the project InnOPlan, funded by the German Federal Ministry for Economic Affairs and Energy (BMWi, FKZ 01MD15002).

References

Armony, M., S. Israelit, A. Mandelbaum, Y. N. Marmor, Y. Tseytlin and G. B. Yom-Tov (2015). “On patient flow in hospitals: A data-based queueing-science perspective.” *Stochastic Systems* 5 (1), 146–194.

Breiman, L. (2001). “Random forests.” *Machine Learning* 45 (1), 5–32.

Cardoen, B., E. Demeulemeester and J. Beliën (2010). “Operating room planning and scheduling: A literature review.” *European Journal of Operational Research* 201 (3), 921–932.

Chandrashekar, G. and F. Sahin (2014). “A survey on feature selection methods.” *Computers and Electrical Engineering* 40 (1), 16–28.

Drucker, H., C. J. C. Burges, L. Kaufman, A. Smola and V. Vapnik (1997). “Support vector regression machines.” *Neural Information Processing Systems* 9, 155–161.

Eijkemans, M. J. C. J. C., M. van van Houdenhoven, T. Nguyen, E. Boersma, E. W. W. Steyerberg and G. Kazemier (2010). “Predicting the Unpredictable A New Prediction Model for Operating Room Times Using Individual Characteristics and the Surgeon’s Estimate.” *The Journal of the American Society of Anesthesiologists* 112 (1), 41–49.

Friedman, J., T. Hastie and R. Tibshirani (2010). “Regularization Paths for Generalized Linear Models via Coordinate Descent.” *Journal of Statistical Software* 33 (1), 1–22.

Gomes, C., B. Almada-Lobo, J. Borges and C. Soares (2012). "Integrating Data Mining and Optimization Techniques on Surgery Scheduling." In: *Proceedings of the 8th International Conference on Advanced Data Mining and Applications*, 589–602.

Guyon, I. and A. Elisseeff (2003). "An Introduction to Variable and Feature Selection." *Journal of Machine Learning Research* 3, 1157–1182.

Hastie, T., R. Tibshirani and J. Friedman (2009). *The elements of statistical learning*. Springer-Verlag.

Hosseini, N., M. Y. Sir, C. J. Jankowski and K. S. Pasupathy (2015). "Surgical Duration Estimation via Data Mining and Predictive Modeling: A Case Study." In: *AMIA Annual Symposium Proceedings Vol. 2015*, 640–648.

Jebali, A. and A. Diabat (2015). "A stochastic model for operating room planning under capacity constraints." *International Journal of Production Research* 53 (24), 7252–7270.

Kayış, E., T. T. Khaniyev, J. Suermondt and K. Sylvester (2015). "A robust estimation model for surgery durations with temporal, operational, and surgery team effects." *Health Care Management Science* 18 (3), 222–233.

Kayış, E., H. Wang, M. Patel, T. Gonzalez, S. Jain, R. J. Ramamurthi, C. Santos, S. Singhal, J. Suermondt and K. Sylvester (2012). "Improving prediction of surgery duration using operational and temporal factors." In: *AMIA Annual Symposium Proceedings Vol. 2012*, 456–62.

Kohavi, R. (1995). "A Study of Cross-Validation and Bootstrap for Accuracy Estimation and Model Selection." In: *Proceedings of the 14th International Joint Conference on Artificial Intelligence*, 1137–1143.

Lamiri, M., X. Xie, A. Dolgui and F. Grimaud (2008). "A stochastic model for operating room planning with elective and emergency demand for surgery." *European Journal of Operational Research* 185 (3), 1026–1037.

Li, Y., S. Zhang, R. F. Baugh and J. Z. Huang (2009). "Predicting surgical case durations using ill-conditioned CPT code matrix." *IIE Transactions* 42 (2), 121–135.

Lowndes, B., C. A. Thiels, E. B. Habermann, J. Bingener, S. Hallbeck and D. Yu (2016). "Impact of patient factors on operative duration during laparoscopic cholecystectomy: evaluation from the National Surgical Quality Improvement Program database." *The American Journal of Surgery* 212 (2), 289–296.

Macario, A. (2010). "What does one minute of operating room time cost?" *Journal of Clinical Anesthesia* 22 (4), 233–236.

Master, N., D. Scheinker and N. Bambos (2016). "Predicting Pediatric Surgical Durations.".

Meier, J., A. Dietz, A. Boehm and T. Neumuth (2015). "Predicting treatment process steps from events." *Journal of Biomedical Informatics* 53, 308–319.

Oshiro, T. M., P. S. Perez and J. A. Baranauskas (2016). "How Many Trees in a Random Forest?" In: P. Perner (Ed.), *Machine Learning and Data Mining in Pattern Recognition*. Cham: Springer International Publishing, Vol. 9729, 58–72.

Pedregosa, F., O. Grisel, R. Weiss, A. Passos, M. Brucher, G. Varoquax, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, … M. Brucher (2011). "Scikit-learn: Machine Learning in Python." *Journal of Machine Learning Research* 12, 2825–2830.

Polato, M., A. Sperduti, A. Burattin and M. de Leoni (2014). "Data-aware remaining time prediction of business process instances." In: *Proceedings of the 2014 International Joint Conference on Neural Networks*, 816–823.

Polato, M., A. Sperduti, A. Burattin and M. de Leoni (2016). "Time and Activity Sequence Prediction of Business Process Instances." *arXiv Preprint arXiv:1602.07566*.

Röder, N., M. Wiesche, M. Schermann and H. Krcmar (2015). "Workaround Aware Business Process Modeling." In: *Proceedings of the 12th International Conference on Wirtschaftsinformatik (WI 2015)*, 482–496.

Saadouli, H., B. Jerbi, A. Dammak, L. Masmoudi and A. Bouaziz (2015). "A stochastic optimization and simulation approach for scheduling operating rooms and recovery beds in an orthopedic surgery department." *Computers & Industrial Engineering* 80, 72–79.

Schultz, K., P. Carayon, A. S. Hundt and S. R. Springman (2007). "Care transitions in the outpatient surgery preoperative process: Facilitators and obstacles to information flow and their consequences." *Cognition, Technology and Work* 9 (4), 219–231.

ShahabiKargar, Z., S. Khanna, N. Good, A. Sattar, J. Lind and J. O'Dwyer (2014). "Predicting Procedure Duration to Improve Scheduling of Elective Surgery." In: *Proceedings of the Pacific Rim International Conference on Artificial Intelligence*, 998–1009.

Smola, a J. and B. Scholkopf (2004). "A tutorial on support vector regression." *Statistics and Computing* 14, 199–222.

Stepaniak, P. S., C. Heij and G. De Vries (2010). "Modeling and prediction of surgical procedure times." *Statistica Neerlandica* 64 (1), 1–18.

Thiels, C. A., D. Yu, A. M. Abdelrahman, E. B. Habermann, S. Hallbeck, K. S. Pasupathy and J. Bingener (2017). "The use of patient factors to improve the prediction of operative duration using laparoscopic cholecystectomy." *Surgical Endoscopy* 31 (1), 333–340.

Tucker, A. L., W. S. Heisler and L. D. Janisse (2014). "Designed for Workarounds: A Qualitative Study of the Causes of Operational Failures in Hospitals." *The Permanente Journal* 18 (3), 33–41.

Van der Aalst, W. M. P., M. H. Schonenberg and M. Song (2011). "Time prediction based on process mining." *Information Systems* 36 (2), 450–475.

van Dongen, B. F., R. A. Crooy and W. M. P. van der Aalst (2008). "Cycle Time Prediction: When Will This Case Finally Be Finished?" In: *Proceedings of the 16th International Conference on Cooperative Information Systems, CoopIS 2008, OTM 2008. Part I, Lecture Notes in Computer Science (Vol. 5331)*, 319–336.

Wolpert, D. H. (1996). "The Lack of A Priori Distinctions Between Learning Algorithms." *Neural Computation* 8 (7), 1341–1390.