8-2019

# Data-driven surgical duration prediction model for surgery scheduling: A case-study for a practice-feasible model in a public hospital

Kar Way TAN
*Singapore Management University*, kwtan@smu.edu.sg

Francis Ngoc Hoang Long NGUYEN
*Singapore Health Services*

Boon Yew ANG
*Singapore Health Services*

Jerald GAN
*Singapore Management University*, tcgan.2015@sis.smu.edu.sg

Sean Shao Wei LAM
*Singapore Management University*, seanlam@smu.edu.sg



## Citation

TAN, Kar Way; NGUYEN, Francis Ngoc Hoang Long; ANG, Boon Yew; GAN, Jerald; and LAM, Sean Shao Wei. Data-driven surgical duration prediction model for surgery scheduling: A case-study for a practice-feasible model in a public hospital. (2019). *2019 IEEE 15th International Conference on Automation Science and Engineering (CASE): Vancouver, Canada, August 22-26: Proceedings*. 275-280.
**Available at:** https://ink.library.smu.edu.sg/sis_research/4688

# Data-Driven Surgical Duration Prediction Model for Surgery Scheduling: A Case-Study for a Practice-Feasible Model in a Public Hospital

Kar Way Tan[1], Francis Ngoc Hoang Long Nguyen[2], Boon Yew Ang [2], Jerald Gan[1] and Sean Shao Wei Lam[3]

*Abstract*— **Hospitals have been trying to improve the utilization of operating rooms as it affects patient satisfaction, surgery throughput, revenues and costs. Surgical prediction model which uses post-surgery data often requires high-dimensional data and contains key predictors such as surgical team factors which may not be available during the surgical listing process. Our study considers a two-step data-mining model which provides a practical, feasible and parsimonious surgical duration prediction. Our model first leverages on domain knowledge to provide estimate of the first surgeon rank (a key predicting attribute) which is unavailable during the listing process, then uses this predicted attribute and other predictors such as surgical team, patient, temporal and operational factors in a tree-based model for predicting surgical durations. Experimental results show that the proposed two-step model is more parsimonious and outperforms existing moving averages method used by the hospital. Our model bridges the *research-to-practice* gap by combining data analytics with expert's inputs to develop a deployable surgical duration prediction model for a real-world public hospital.**

## I. INTRODUCTION

Operating rooms (ORs) are among the highest expenditure items in hospitals and scheduling them is a complex and multi-faceted problem. Surgery duration is affected by factors from multiple areas such as different medical disciplines, experiences of the surgical team and patient's health conditions. Underestimation of surgery duration leads to downstream cases being delayed or canceled and incurs additional unplanned cost of overtime work. Overestimation of surgery duration produces lower utilization and throughput of the ORs. Inaccurate surgical scheduling results in operational and economic impacts on hospital operations.

At the point of the study, the Study Hospital (SH) used a simple Moving Average (MA) to predict surgical duration. The moving average [1] is commonly employed by many commercial software supporting the operation of hospitals. MA deployed in SH provides an estimate based on the last 5 historical records of the same procedure code (a standardized set of procedures regulated by the national regulatory body) and surgeon. If there is insufficient retrospective cases, the system selects last 5 using only the procedure code. The estimate can be modified by user's input in conjunction with the advice from physician in-charge of the patient. The MA prediction is not ideal in capturing the fluctuations in surgical duration due to variation pertaining to any surgery. In some operations, the physician who provided the estimate may not be the lead surgeon conducting the actual surgery.

Going beyond the commercial method of using moving average as a prediction, existing literature offers alternatives. Some of which include predicting surgical duration based on patient factors, procedure complexity, surgeon or operating team factors, operational and temporal factors. We found multiple challenges in applying the models to our context at SH, which is a large public hospital. For example, the surgeon-specific factors such as individual surgeon result in large number of categorical variables. In addition, in the hospital of interest, the subsidized patients are not offered the option to select the operating surgeon, thus the first surgeon is not known until shortly before the day of surgery. Hence, key predictors such as the information about the surgeon history and surgeon work rate are not available during the Listing Process – a process which allocates physical resources (e.g., OR) for the surgery.

We propose and contribute a two-phase tree-based classification and ensemble prediction model. The first phase involves feature engineering for predicting the surgeon rank (instead of features of the individual surgeon) by combining operational domain knowledge and a classification method. The second phase encompasses an ensemble approach using Gradient Boosting Machine (GBM). Both techniques are tree-based methodologies to deal with substantial number of categorical attributes. Our modeling approach incorporates key predictors such as patient history, complexity of the surgery, discipline, surgeon rank and temporal factors such as moving averages. We combined the advantages of using different predictors while ensuring a parsimonious yet practice-feasible prediction model for the operational requirement of a large public hospital.

The baseline model of using all predictors (including surgical team, surgical complexity, patient factors, temporal factors) has 10% improvement over the Moving Average model. Our final model which included only the rank of the first surgeon and three other groups of factors (surgical complexity, patient and temporal factors) surpasses the other models in terms of prediction performance, and it is viable

[1]Kar Way Tan and Jerald Gan are with School of Information Systems, Singapore Management University, kwtan@smu.edu.sg, tcgan.2015@sis.smu.edu.sg

[2]Ngoc Hoang Long Nguyen and Boon Yew Ang are with the Health Services Research Centre at Singapore Health Services, {ang.boon.yew, nguyen.ngoc.hoang.long}@singhealth.com.sg

[3]Sean Shao Wei Lam is with Lee Kong Chian School of Business, Singapore Management University, Health Services and Systems Research, Duke NUS Medical School and Health Services Research Centre at Singapore Health Services, seanlam@smu.edu.sg, gmslasws@nus.edu.sg

for both private and subsidized patients at the public hospital.

The rest of the paper is structured as follows: Section II provides a literature review on related works. Section III details the background descriptions of the study hospital, dataset and the business processes supporting the surgery listing process. Section IV states the objective and hypotheses of our study. Section V describes our proposed approach to the practice-feasible prediction model. Section VI shows the results of our study. Finally, we draw our conclusion in VII.

## II. Literature Review

Among the literature on surgery prediction, one approach is the distribution method by Stepaniak et al. [2] where the actual surgery duration data was plotted on histograms for distribution fitting. The authors found that the prediction of duration can be improved by incorporating surgeon factors such as surgeon work rate. The work rate of each individual surgeon is one of the key predictors for surgical duration.

Other approaches include more predictors such as patient factors, case-related and procedure complexity, surgeon or operating team factors. Strum et al. [3] proposed a linear regression model based on surgeon, type of anesthesia, ICD-9 codes, American Society of Anesthesiologist (ASA) status, age and gender of the patient as predictors for each Common Procedural Terminology (CPT). Kayis et al. [4] first presented a model using operational and temporal factors and found that the operational factors are promising in improving the predictability of surgery duration. It was also reported that variability in such estimates remained high, necessitating caution in using them when optimizing OR schedules. In another work by Kayis et al. [5], authors presented a duration adjustment method based on combination of operational, surgical team and temporal factors. The study showed that medical staff and team experience related factors could further improve the surgery duration estimates. Eijkenmans et al. [6] used a combination of operational, patient and surgical team characteristics to estimate surgical duration. The result demonstrated that potential strong predictive factors were surgical team and patient's health factors such as number and experience of surgeons and anesthesiologists, patient's age and sex and number of previous hospital admissions.

Among the literature which specialty of surgery was considered, Hosseini et al. [7] proposed a hybrid method in which first step consisted of factor analysis and classification model to reduce the number of categories of procedural codes followed by a stepwise regression method to predict surgical duration based on priority class, procedure category, ASA class, age, patient class and specialty. The result asserted that few specialties such as orthopedics surgeries and surgical oncology attained better predictions than others. In a more specific study, Shahabikargar et al. [8] applied ensemble algorithms (e.g., Random Forest, Bagging and LSBoost) prior to a regression model to predict surgical duration at specialty level, e.g., Cardio-horacic, ENT, Gynaecology, Neurosurgery and few others. The performance of prediction model varied significantly across different specialties but showed more potential for Ophthalmology, Neurosurgery and Vascular surgeries.

In the broader aspect, Rickert et al. [9] approached estimation and optimization problem from the business process management perspective. The surgery duration predictions were based on accessing the process length of surrounding stations to better assess the status of the operating room for more specific forecast.

Surgery, patient, surgical team and temporal factors are good potential predictors which we apply in our model. However, to fit specific needs in our context in a large public hospital with more than forty disciplines, some of these methods are not directly applicable. For example, surgeon-specific factors will not be efficient as they result in large number of categorical data. Linear regression is not well-suited for our dataset with high-dimensional categorical variables such as procedure codes and disciplines. Finally, we needed a method to handle surgeries for subsidized patients in which the surgical team is not known during the listing process.

## III. Case Scenario

### A. Process

Our study involves a large public hospital in Asia. The process of listing a surgery for scheduling is depicted in Fig. 1. The process comprises of two sub-processes, i.e., surgery ordering by consultant in-charge and surgery listing by a listing nurse.

When surgery is requested, the attending physician makes an *order* for an operation. Based on the prognosis of the patient's symptoms and conditions, the surgeon chooses procedure code(s) to represent the operation(s) required. At the same time, based on preliminary results (e.g., blood test) and assessment, an ASA health score (ranging from 1 to 6) which represents health status is recorded for the patient. Subsequently, the patient health statuses and surgery requirements are entered into the system. The consultant may also indicate an estimated duration of the surgery based on experience which is meant for the next step. This set of information is then combined and be passed to a listing personnel(typically a nurse).

The listing personnel consolidates the information provided by the medical consultant and the surgery duration estimate provided by the system based on the Moving Average (MA) to list the surgery by selecting the physical resources required within a group of ORs. There are multiple factors preventing accurate surgical listing duration. Firstly, about 45% of the surgery orders do not come with an estimated duration. Secondly, about 40% of the patients handled in the public hospital are subsidized patients and hence the principal surgeon may not be the physician who made the initial surgery order. To further complicate matters, the listing process is non-trivial due to a set of policies governing the use of ORs, e.g., disciplinary ORs can only be made available to the specific discipline until a specific window when they can be opened for surgeries in the other disciplines. This results in differences in operational requirement for setting up and cleaning the OR for surgery outside the designated discipline.

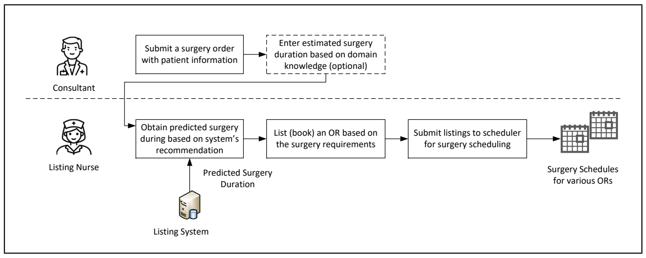


Fig. 1. Process of Ordering and Listing of Surgery

### B. Data

In this data analytics project, we work with $41,000$ surgical records involving 41 medical disciplines and 33 data attributes. The selected data incorporated only the single-procedure elective surgeries collected in 2016 and 2017. The more complex multi-procedure surgeries involve slightly different listing and scheduling process described and hence omitted from our study. The data is then split into 75% for training our models and 25% for verification.

The variables found in the data can be generalized into four broad factors which are associated with the surgical duration based on previous studies in the literature – Surgery Factors, Patient's Factors, Surgical Medical Team Factors and Temporal Factors. The list of elements is listed in Table I where categorical data is set as *C*, numerical data as *N* and unique identifiers as *ID*. For each categorical data, we show the number of categories found in the data and the mean count for each category. The number of categories is not applicable (N.A.) for numeric and unique identifier variables.

The Surgical Team Factors outlined the experience and composition of the surgery crew which affect the surgery duration. For example, if there is a student participating in the surgery, the surgery might potentially take longer than expected. However, there are also contrary results: in our preliminary investigation, we found a weak negative correlation between the presence of medical student and the surgical duration. Further investigation with domain experts subsequently showed that the first surgeon is the key influencing factor and medical student is only present when the first surgeon is at least of a senior rank.

### C. Research-to-Practice Gap

A prior phase of our project and many works reported in the literature use most of the attributes in the dataset. Yet, the main challenge is to overcome the limitation due to data unavailability (members of the surgical team would only be known days before the surgery, and not during the listing process). This is a characteristic unique to subsidized patients in the public hospital, who are operated by any surgical team on duty for the discipline.

TABLE I
TABLE SHOWING THE ATTRIBUTES IN THE DATASET

| Variable | Type | No. of Categories | Mean Count |
|---|---|---|---|
| **Surgery Factors** | | | |
| Department Code | C | 41 | 1018.51 |
| Op. Theatre Code (OT) | C | 38 | 1098.92 |
| OT Location Code | C | 3 | 13919.67 |
| Proc. Code (PC) | C | 1308 | 31.93 |
| Proc. Desc. | C | 1331 | 8.1 |
| Proc. Surgical Table Code (TC) | C | 22 | 1898.14 |
| Operation Risk | C | 4 | 10439.72 |
| Type Of Anaesthesia | C | 21 | 1988.51 |
| Method Of Operation | C | 5 | 8351 |
| **Patient Factors** | | | |
| Type Of Operation | C | 2 | 20879.45 |
| Gender | C | 2 | 20879.45 |
| Race | C | 20 | 2087.95 |
| Weekday | C | 6 | 5965.57 |
| ASA Status | N | N.A. | N.A. |
| **Surgical Team Factors** | | | |
| Surgical Team Size | N | N.A. | N.A. |
| 1st Surgeon ID | ID | 717 | 61.84 |
| 2nd Surgeon ID | ID | 812 | 54.61 |
| 3rd Surgeon ID | ID | 736 | 60.25 |
| 1st Surgeon Title | C | 13 | 3410.91 |
| 2nd Surgeon Title | C | 812 | 54.61 |
| 3rd Surgeon Title | C | 13 | 3410.91 |
| P.Anaes. ID | ID | 142 | 312.27 |
| P.Anaes. Title | C | 8 | 5542.74 |
| Asst Anaes. Title | C | 11 | 4031.08 |
| Surgical Student | C | 3 | 14780.64 |
| Anaes. Student | C | 3 | 14780.64 |
| Consultant ID | ID | 334 | 132.76 |
| Consultant Title | C | 11 | 4031.08 |
| **Temporal Factors** | | | |
| Moving Avg Dept | N | N.A. | N.A. |
| Moving Avg TC | N | N.A. | N.A. |
| Moving Avg OT | N | N.A. | N.A. |
| Moving Avg Diag | N | N.A. | N.A. |
| Moving Avg PC | N | N.A. | N.A. |

## IV. PROBLEM DEFINITION

The desired outcome for our proposed surgical duration prediction model is for it to be functional by the listing practitioners and be used in robust surgery scheduling across disciplines. The prediction model must contain only data attributes which are available during listing. With that, we aim to verify the following hypotheses in our investigations for our practice-feasible model for the public hospital with consideration of subsidized patients.

*Hypothesis 1:* At least one of the factors among the Surgical Team Factors is significant in predicting Surgical Duration.

*Hypothesis 2:* The seniority of First Surgeon is a significant (and sufficiently) good predictor in predicting Surgical Duration.

## V. APPROACH

Due to a large number of categorical data pertaining in the dataset for prediction, we propose the use of a tree-based model, in particular, we chose Gradient Boosting Machines (GBM) as the learning model.

### A. Gradient Boosting Machine

GBM is a class of ensemble algorithm which builds a prediction tree by combining a set of smaller and weaker predictive trees. The algorithm employs the logic where predictors learn from the previous predictors to improve[10]. This is accomplished through the building of the trees sequentially. GBM is selected as a suitable learning model for our use case study compared to conventional decision trees because GBM allows the bias and variance to be reduced through ensemble and the model is able to better handle categorical variables.

### B. Definition of Models

A summary of the predictive models based on GBM is listed in Table II and Fig. 2.

TABLE II

TABLE SHOWING THE VARIOUS MODELS USED IN OUR ANALYSIS

| Model | Description |
|---|---|
| MA | Moving Average and estimation from domain expert |
| $M_0$ | Baseline model using all 4 broad factors |
| $M_1$ | Practice-feasible model using Surgery, Patient and Temporal Factors but without the 12 unknown Surgical Team Factors |
| $M_2$ | Practice-feasible model using factors in $M_1$ + First Surgeon's Rank prediction |

Our first generation predictive model $M_0$ considered all the listed variables in Table I. $M_0$ becomes our baseline model for comparison among the updated predictive models. $M_0$ is also a good starting point in our analysis as the attribute selection analysis provided us with insights on the significant variables contributing to the prediction model.

Model $M_1$ is derived by removing 12 variables among the Surgery Team factors which are unknown during the listing process. The missing factors represent about 36% of the variables used in $M_0$.

Based on $M_0$, we ran a Variable Importance test in which Moving Averages, OT location, Team Size and First Surgeon ID were among the top 10 predictors. However, team size and first surgeon remained unknown during the listing process. Based on domain knowledge, we understand that for non-subsidized (using private ward class) patients, the Consultant-in-charge will most likely to be the First Surgeon performing the surgery. The First Surgeon's rank is a hierarchical variable that includes 7 main categories (a few smaller sub-categories have been aggregated for this analysis), namely Medical officer (MO), Resident (Res), Senior Resident (SRES), Registrar (Reg), Associate (AC), Consultant (Con) and Senior Consultant (Scon).

Using this information, we designed a new model $M_2$ by building another predictive model to first determine the seniority of First Surgeon. The predicted First Surgeon is then added as an input attribute to model $M_1$ to form model $M_2$. The rank of First Surgeon is more robust than predicting individual surgeon ID or work rates as individual predictions involve too many categories and not applicable to new surgeons. Illustrated in Fig. 2, the prediction of First Surgeon Rank requires additional data – the patient's ward class. Typically, patients who opt for ward class A1, A2 and B1 are private patients (non-subsidized) while subsidized patients can only choose ward class B2 and C. By using the additional data based on ward class and the domain understanding that private patient would typically use his/her consultant-in-charge as the surgeon, we constructed a separate prediction model based on Decision Tree (using CART algorithm) to predict the rank of First Surgeon.

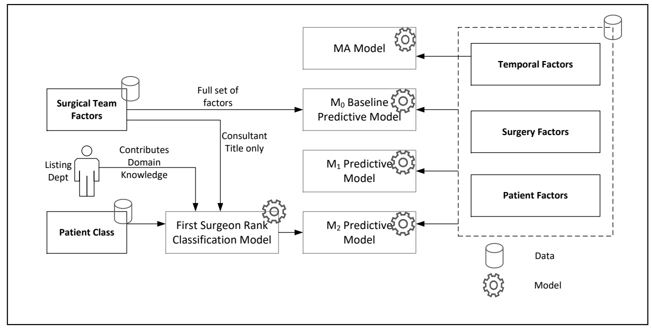


Fig. 2. Summary of the various predictive models and the dataset used

### C. First Surgeon Prediction Model

We hypothesize that the First Surgeon's rank is correlated with the complexity of the surgery, which characteristics can be observed from the Surgery Factors and Patient Factors. In short, when a surgery is complex and the patient is of high risk, the rank of the first surgeon is higher.

Hence, we proposed using Decision Tree algorithm, in particular, RPART package in R, for prediction of First Surgeon Rank. We begin with training the model for subsidized patients using 7 categories of surgeon ranks. The accuracy of the tree reported was 71.35%. We noticed that the differentiating ranks which make significant differences in the surgical durations are the top 3 ranks, i.e., Associate (AC), Consultant (Con) and Senior Consultant (Scon). We then simplified the decision tree to consider only 4 categories, namely Associate, Consultant and Senior Consultant and "Registrar and below'. The accuracy of the model improved to over 90%.

By using our test dataset with approximately 60% private patients and 40% subsidized patients, the predictive accuracy of the First Surgeon Rank in conjunction with both domain knowledge and decision tree prediction was found to be 95.02%. This is a very strong prediction providing us with the ability to add First Surgeon Rank into the surgery prediction model. The algorithm table in III shows the detailed steps for deriving model $M_2$.

## VI. EXPERIMENTAL RESULTS

### A. Performance Metrics

To compare the results across the models, we have considered standard measurements of errors such as Root Mean Square Errors (RMSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). However, we have

TABLE III

$M_2$ ALGORITHM WHICH INCLUDES PREDICTION OF FIRST SURGEON'S RANK

| | *Steps* |
|---|---|
| 01. | Filter the train dataset by patients who opted for subsidized class wards. |
| 02. | Train the FirstSurgeonRankDecisionTree model using Surgery Factors and Patient Factors in the dataset. |
| 03. | Given a new patient to be listed for surgery, |
| 04. | IF patient opted for Private class wards |
| 05. | First Surgeon Rank $\leftarrow$ Consultant-In-Charge Rank (based on 4 categories) derived from Consultant's Title |
| 06. | ELSE |
| 07. | First Surgeon Rank $\leftarrow$ FirstSurgeonRankDecisionTree(Surgery Factors, Patient Factors) |
| 08. | ENDIF. |
| 09. | Add First Surgeon Rank to dataset with Surgery, Patient and Moving Average Factors. Use $M_2$ model to predict surgical duration. |
| 10. | Return the surgical duration prediction for a new surgery. |

finally chosen RMSE as our main final analysis metric as we wanted to penalize large errors by taking the square of errors. In our initial investigation, $M_0$ showed 10% improvements in RMSE over MA. In Table IV, we show the RMSE of the various GBM prediction models.

TABLE IV

RESULTS FOR VARIOUS MODELS IN OUR ANALYSIS

| **Model** | **RMSE** |
|---|---|
| $M_0$ | 55.37 |
| $M_1$ | 58.28 |
| $M_2$ | 54.95 |

$M_0$ has better RMSE scores compared to $M_1$ that indicates our Hypothesis 1 may be true, i.e., removal the Surgical Team Factors has affected the performance of the model as the factors are significant predictors of surgical duration. $M_2$ has better RMSE compared to $M_1$ giving us an indication that Hypothesis 2, i.e., First Surgeon Rank, representing the surgery team formation is a significant contributor towards the dependent variable.

### B. *Statistical Diagnosis on Pairwise Errors Across Models*

To evaluate if the results are indeed statistically significant, we used pair-wise comparisons for each test data point. In deciding between parametric and non-parametric test, we ran the Shapiro-Wilk test of normality on the errors produced by the models and also examined the Normal Q-Q plots (See Fig. 3). The Q-Q plot suggests that the errors are not following normal distribution. The p-values of the normality tests are below the alpha value of 0.05. Therefore, the null hypothesis of normality test is rejected as there is little evidence that the errors are normally distributed. Thus, non-parametric one-tailed Wilcoxon Signed Ranked Test (with $\alpha = 0.05$) is chosen for the pairwise comparison of error terms produced by any two models. This analysis also supported our choice of using tree-based training model which does not require a normality assumption on the error distribution.

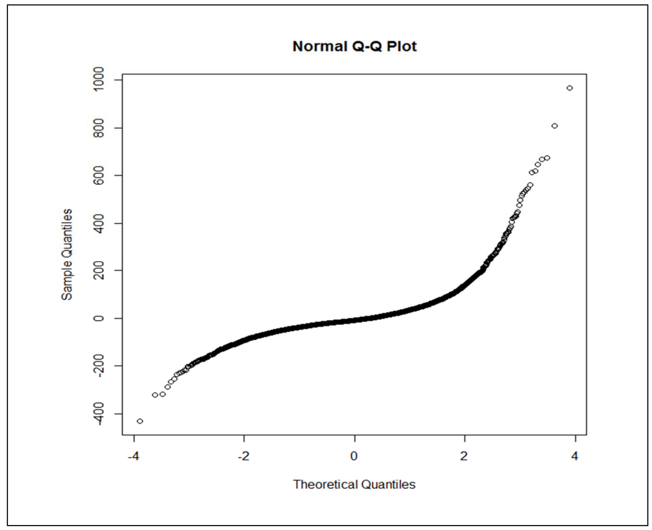


Fig. 3. Quantile-Quantile Plot for Investigating Normality

Firstly, we ran the Wilcoxon Signed Ranked Test between models $M_0$ and $M_1$:

TABLE V

WILCOXON TEST FOR HYPOTHESIS 1

| | |
|---|---|
| $H_0$: | No significant differences between the errors of $M_0$ and $M_1$ |
| $H_1$: | $M_1$ has higher errors than $M_0$ |
| p-value: | 0.0002552 $\rightarrow$ Reject $H_0$ |

The rejection of $H_0$ indicates there is statistically sufficient evidence that $M_1$ has higher errors than $M_0$ in terms of RMSE and removing Surgery Team Factors resulted in poorer prediction of surgical duration. This supports Hypothesis 1 where at least one of the factors among the Surgical Team Factors is significant in predicting the surgical duration.

Next, we ran the Wilcoxon Signed Ranked Test between models $M_0$ and $M_2$:

TABLE VI

WILCOXON TEST FOR HYPOTHESIS 2

| | |
|---|---|
| $H_0$: | No significant differences between the errors of $M_0$ and $M_2$ |
| $H_1$: | $M_0$ has higher errors than $M_2$ |
| p-value: | 0.001869 $\rightarrow$ Reject $H_0$ |

Similarly, Wilcoxon test for Hypothesis 2 reveals $M_0$ has statistically significant higher errors than $M_2$ in terms of RMSE, where $M_2$ has the seniority of First Surgeon is added to the prediction model. This supports Hypothesis 2 where the Rank of First Surgeon is an important predictor for the surgical duration, resulting in better RMSE in model $M_2$ as the model becomes more parsimonious than $M_0$. The Rank of First Surgeon is sufficient in replacing the Surgical Team Factors which are unavailable during the listing process.

### C. Further Discussions

We recognize that the performance of the model is dependent on the prediction results of the First Surgeon's rank and its accuracy is attributed to the nearly-balanced proportions of private and subsidized patients. However, we believe that the findings we gathered from this project can be applied to other projects in two aspects:

1) Our study may help in simplifying some of the existing models proposed in the literature where surgical team factors were used. Instead of using the entire surgical team factors as predictors, using only the rank of the first surgeon might achieve similar results with a more parsimonious model.
2) The classification model for predicting the First Surgeon Rank may be applicable to another hospital where surgeon is unknown during listing or scheduling of surgeries. The model can also be applied to resource estimation or allocation for scheduling different ranks of surgeons.

Future work may consider performing sensitivity analysis on the proportion of private and subsidized patients and evaluate the impact of the performance of the First Surgeon Rank prediction model to determine if it is dependent and unique to our scenario with nearly-balanced proportions of the two classes. Research could also extend to using the output of the Surgical Duration prediction model in a robust OR scheduling optimization model. By using our proposed surgical duration prediction model to obtain the means and variances of surgeries by discipline types, it can potentially be incorporated into robust scheduling model for each discipline type with minimizing robust cost as the objective function for creating schedules for operating rooms.

## VII. CONCLUSION

In this work, we showed that surgical prediction models using post-surgery data contained key predictors which were unavailable during surgical listing, hence not suitable for real-world implementation in a public hospital. Our final model considered a two-step prediction paradigm which first used a classification method to predict the rank of the first surgeon, then a tree-based model (suitable for high dimensional categorical data) using Gradient Boosting Model to achieve significantly lower root mean square error compared to the baseline model. Our proposed model provides the listing practitioners valuable insight during the listing and scheduling of surgeries. While the model considers other key predictors such as patient factors, surgical type and complexity as well as temporal factors (e.g. moving average), it does not require the surgical team formation which is unavailable during listing for the subsidized patients in a public hospital. Our two-step prediction model provides a practice-feasible, more parsimonious improvement over a baseline model which uses all prediction factors in the surgical duration prediction model.

## ACKNOWLEDGMENT

The authors would like to thank the support of staff members in the Listing Department of our study hospital.

## REFERENCES

[1] A. Macario and F. Dexter, "Estimating the duration of a case when the surgeon has not recently scheduled the procedure at the surgical suite," *Anesthesia and Analgesia*, vol. 89, no. 5, pp. 1241 – 1245, November 1999.

[2] P. S. Stepaniak, C. Heij, and G. de Vries, "Modeling and prediction of surgical procedure times," *Statistica Neerlandica*, vol. 64, no. 1, pp. 1 – 18, 2010.

[3] D. P. Strum, A. R. Sampson, J. H. May, and L. G. Vargas, "Surgeon and type of anesthesia predict variability in surgical procedure times," *Anesthesiology*, vol. 92, no. 5, pp. 1454 – 1466, 2000.

[4] E. Kayis, H. Wang, M. Patel, T. Gonzalez, S. Jain, R. Ramamurthi, C. Santos, S. Singhal, J. Suermondt, and K. Sylvester, "Improving prediction of surgery duration using operational and temporal factors," in *AMIA Annual Symposium Proceedings*. AMIA, 2012, pp. 456–462.

[5] E. Kayış, T. T. Khaniyev, J. Suermondt, and K. Sylvester, "A robust estimation model for surgery durations with temporal, operational, and surgery team effects," *Health Care Management Science*, vol. 18, no. 3, pp. 222–233, Sep 2015. [Online]. Available: https://doi.org/10.1007/s10729-014-9309-8

[6] M. J. Eijkenmans, M. van Houdenhoven, T.Nguyen, E.Boersma, E. Steyerberg, and G. Kazemier, "Predicting the unpredictable: a new prediction model for operating room times using individual characteristics and the surgeon's estimate," *Anesthesiology*, vol. 112, pp. 41 – 49, January 2010.

[7] N. Hosseini, M. Sir, C. Jankowski, and K. Pasupathy, "Surgical duration estimation via data mining and predictive modeling: A case study," in *AMIA Annual Symposium Proceedings*. AMIA, 2015, pp. 640–648.

[8] Z. Shahabikargar, S. Khanna, A. Sattar, and J. Lind, "Improved prediction of procedure duration for elective surgery," *Studies in Health Technology and Informatics*, pp. 133–138, 2017.

[9] M. Ricker, M. Premm, A. Klein, L.Kirilov, H.Kenngott, M. Apitz, M. Wagner, and L. Ternes, "Predicting the duration of surgeries to improve process efficiency in hospitals," *Procs of the 25th European Conference on Information Systems (ECIS)*, pp. 2842 – 2851, June 2017.

[10] P. Grover, "Gradient boosting from scratch - ml review - medium," December 2017, https://medium.com/mlreview/gradient-boosting-from-scratch-1e317ae4587d, Last accessed on 2019-03-10.