Journal of Medical Systems (2019) 43: 32
https://doi.org/10.1007/s10916-018-1151-y

SYSTEMS-LEVEL QUALITY IMPROVEMENT

A Machine Learning Approach to Predicting Case Duration for Robot-Assisted Surgery

Beiqun Zhao[1,2] · Ruth S. Waterman[3] · Richard D. Urman[4] · Rodney A. Gabriel[2,3]

Received: 23 August 2018 / Accepted: 25 December 2018 / Published online: 5 January 2019

Abstract

Robot-assisted surgery (RAS) requires a large capital investment by healthcare organizations. The cost of a robotic unit is fixed, so institutions must maximize use of each unit by utilizing all available operating room block time. One way to increase utilization is to accurately predict case durations. In this study, we sought to use machine learning to develop an accurate predictive model for RAS case duration. We analyzed a random sample of robotic cases at our institution from January 2014 to June 2017. We compared the machine learning models to the baseline model, which is the scheduled case duration (determined by previous case duration averages and surgeon adjustments). Specifically, we used: 1) multivariable linear regression, 2) ridge regression, 3) lasso regression, 4) random forest, 5) boosted regression tree, and 6) neural network. We found that all machine learning models decreased the average root-mean-squared error (RMSE) as compared to the baseline model. The average RMSE was lowest with the boosted regression tree (80.2 min, 95% CI 74.0–86.4), which was significantly lower than the baseline model (100.4 min, 95% CI 90.5–110.3). Using boosted regression tree, we can increase the number of accurately booked cases from 148 to 219 (34.9% to 51.7%, $p < 0.001$). This study shows that using various machine learning approaches can improve the accuracy of RAS case length predictions, which will increase utilization of this limited resource. Further work is needed to operationalize these findings.

Keywords Robot-assisted surgery · Machine learning · OR efficiency · Health economics · Prediction · Case duration

Introduction

Operating rooms (ORs) are responsible for a large proportion of both revenue and cost for most hospitals [1, 2]. Approximately 60% of all patients admitted to the hospital are treated in the OR [3]. Across the country, we increasingly find robotic devices used to assist surgeons during procedures.

This article is part of the Topical Collection on *Systems-Level Quality Improvement*

✉ Beiqun Zhao
beiqunmzhao@gmail.com

1 Department of Surgery, University of California, San Diego, 9300 Campus Point Drive, #7220, La Jolla, CA 92037, USA

2 Department of Medicine, Division of Biomedical Informatics, University of California, San Diego, La Jolla, CA, USA

3 Department of Anesthesiology, University of California, San Diego, La Jolla, CA, USA

4 Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA, USA

Since its introduction into mainstream clinical practice, robot-assisted surgery (RAS) volume has increased annually [4]. RAS has many advantages, but requires large capital investments by hospitals [4, 5]. The costs of the robotic surgical systems are relatively fixed, with prices ranging from $1 million to $2.5 million per unit, plus annual maintenance fees [6]. Therefore, one way for hospitals to maximize revenue is to maximize the use of each robotic unit by utilizing all available robotic surgery block time.

OR efficiency can be increased by more accurate prediction of case lengths [3]. Currently, most hospitals employ two techniques to schedule case duration – surgeon estimates and/or historical case durations. Multiple studies have shown that both of these methods have limited accuracy [7–9]. Because operating room case durations can be affected by many factors, some studies have looked to improve the accuracy of these predictions by using simple regression techniques [3, 10, 11]. However, advanced predictive analytics, such as machine learning, can potentially provide even more accurate estimations. Machine learning utilizes statistical techniques to enable computer systems to "learn" from available data and make accurate predictions on outcomes of interest.

While other industries, such as finance and advertising, have harnessed the power of machine learning, healthcare has been slow to utilize this powerful tool. In this study, we incorporated patient factors, disease factors, and operation settings to construct a predictive model using machine learning algorithms to improve the accuracy of predicted case lengths for a variety of RAS procedures.

Material and methods

After approval by the University of California, San Diego Institutional Review Board, we analyzed a random sample of 500 elective RAS cases at our institution from January 1, 2014 to June 30, 2017. All surgical subspecialties were included in this cohort. All robotic cases were identified by a combination of the term "robotic" or "robot" in the procedure name and by the robotic surgery Common Procedure Terminology code "S2900". We then performed a manual chart review of all cases to ensure that a robotic procedure was performed. We excluded procedures with less than ten instances. Procedures were grouped into 12 categories: bowel resection, cystectomy, low anterior resection, esophageal myotomy, partial nephrectomy, radical nephrectomy, radical prostatectomy, salpingo-oophrectomy, simple hysterectomy, total abdominal hysterectomy and bilateral salpingo-oophrectomy, trans-oral surgery, and ureteral reimplantation. We selected 28 variables (Table 1) to build our predictive models. These variables were selected because of their clinical significance and to minimize collinearity in model-building. A number of these variables have also been used by previous studies to predict case duration in other procedures [3, 8, 10–13].

Statistical analysis

R, a software environment for statistical computing (R version 3.3.2), was used to perform all statistical analysis. We developed various predictive models for case duration, which was defined as the time the patient exited the OR minus the time the patient entered the OR. First, we constructed a baseline model, which tested the ability of scheduled case duration to predict actual case duration by single variable linear regression. Scheduled case duration was the duration of time the case was originally booked for, based on historical averages of case durations and/or surgeon preference. This baseline model served as a reference group for comparison to other predictive models. We then tested multiple machine learning techniques (multivariable linear regression, ridge regression, lasso regression, random forest, boosted regression tree, and neural network) and compared them to the baseline model.

The first machine learning model we tested was multivariable linear regression (MLR). We included all 28 variables and performed no model selection. We then used all 28 variables and performed both ridge regression (RR) and lasso regression (LR). Briefly, ridge regression is similar to maximum likelihood estimation, except a shrinkage penalty is applied to each regression coefficient, and thus, it shrinks the coefficients towards zero. The shrinkage penalty is weighted based on a tuning parameter λ. This provides an advantage in

Table 1 List of the 28 variables used for model building and possible values for each predictor

Predictors	Values	Predictors	Values
Scheduled duration	120–360	OSA	Yes/No
Procedure group	See Table 4	CAD	Yes/No
Elderly (age > 65)	Yes/No	Diabetes mellitus	Yes/No
Obese (BMI >30)	Yes/No	Cirrhosis	Yes/No
Gender	Male/Female	COPD	Yes/No
Combined case	Yes/No	CKD	Yes/No
Robot model	S or Si/Xi/Unknown	ASA classification	1–6
Malignancy	Yes/No	Month of year	January – December
Tumor chest	Yes/No	Time of day	First Start/AM/PM
Tumor abdomen	Yes/No	Day of the week	Monday – Friday/ Weekend
Tumor head and neck	Yes/No		
Tumor pelvis	Yes/No	Anesthesia provider	CRNA/Resident/Attending Only
Tumor retroperitoneum	Yes/No		
Hypertension	Yes/No	Surgery resident PGY	1–5/Fellow/Unknown/ Attending Only
Smoking history	Yes/No		
Atrial fibrillation	Yes/No		

ASA American Society of Anesthesiologists, *BMI* body mass index, *CAD* coronary artery disease, *CKD* chronic kidney disease, *COPD* chronic obstructive pulmonary disease, *CRNA* certified registered nurse anesthetist, *PGY* post-graduate year, *OSA* obstructive sleep apnea

Table 2 Demographics and descriptives for variables used to construct the machine learning models

Variable	Descriptive
Scheduled duration minutes	300
Procedure	
- Bowel resection	22
- Cystectomy	28
- Low anterior resection	50
- Myotomy	14
- Partial nephrectomy	53
- Radical nephrectomy	14
- Radical prostatectomy	124
- Salpingo-oophorectomy (benign)	22
- Simple hysterectomy (benign)	13
- TAH-BSO (malignant)	60
- Trans-oral robotic surgery	13
- Ureteral reimplantation	11
Elderly patient (>65 years-old)	163 (38.4%)
Obese patient (BMI > 30)	145 (34.2%)
Female gender	178 (42.0%)
Combined case	23 (5.4%)
Starting anesthesiologist	
- CRNA	167 (39.4%)
- Resident	179 (42.2%)
- Attending only	78 (18.4%)
Surgery resident post-graduate year	
- 2	4 (0.9%)
- 3	28 (6.6%)
- 4	111 (26.2%)
- 5	68 (16.0%)
- Fellow	139 (32.8%)
- Unknown	72 (17.0%)
- No residents	2 (0.5%)
Robot type	
- S/Si	147 (34.7%)
- Xi	154 (36.3%)
- Unknown	123 (29.0%)
Malignancy	342 (80.7%)
Tumor	
- Chest	1 (0.2%)
- Abdomen	1(0.2%)
- Retroperitoneum	73 (17.2%)
- Head and neck	11 (2.6%)
- Pelvis	260 (61.3%)
Hypertension	160 (37.7%)
Smoking history	17 (4.0%)
Atrial fibrillation	27 (6.4%)
Obstructive sleep apnea	41 (9.7%)
Coronary artery disease	28 (6.6%)
Diabetes mellitus	60 (14.2%)
Cirrhosis	11 (2.6%)
Chronic obstructive pulmonary disease	30 (7.1%)
Chronic kidney disease	34 (8.0%)
ASA classification	
- 1	11 (2.6%)
- 2	165 (38.9%)
- 3	236 (55.7%)
- 4	12 (2.8%)
Day of the week	
- Monday	120 (28.3%)
- Tuesday	81 (19.1%)
- Wednesday	61 (14.4%)
- Thursday	121 (28.5%)
- Friday	39 (9.2%)
- Weekend	2 (0.5%)
Time of the day	
- First start	223 (52.6%)
- AM	127 (30.0%)
- PM	74 (17.4%)
Month of the year	
- January	24 (5.7%)
- February	27 (6.4%)
- March	39 (9.2%)
- April	46 (10.8%)
- May	36 (8.5%)
- June	66 (15.6%)
- July	24 (5.7%)
- August	32 (7.5%)
- September	28 (6.6%)
- October	28 (6.6%)
- November	41 (9.7%)
- December	33 (7.8%)

ASA American Society of Anesthesiologists, *BMI* body mass index, *CRNA* certified registered nurse anesthetist, *TAH-BSO* Total Abdominal Hysterectomy and Bilateral Salpingo-oophrectomy

the bias-variance tradeoff, in which increases in λ may lead to decreased variance but increased bias. Ridge regression uses all predictor variables in the final model, whereby no coefficients will be equal to exactly zero. In contrast, while lasso has very similar properties to ridge regression, it allows smaller, more parsimonious models (i.e., some regression coefficients may equal zero). For both ridge and lasso regressions, the optimal λ value was determined via cross-validation and was used for the final models. For random forest (RF), the number of predictors chosen at each split was set to equal the square root of the total number of predictors and the total number of trees was set at 500. For the boosted regression tree (BRT), the number of trees was set at 10,000, with the shrinkage value set at 0.001. Lastly, for the neural network (NN): 1) dummy

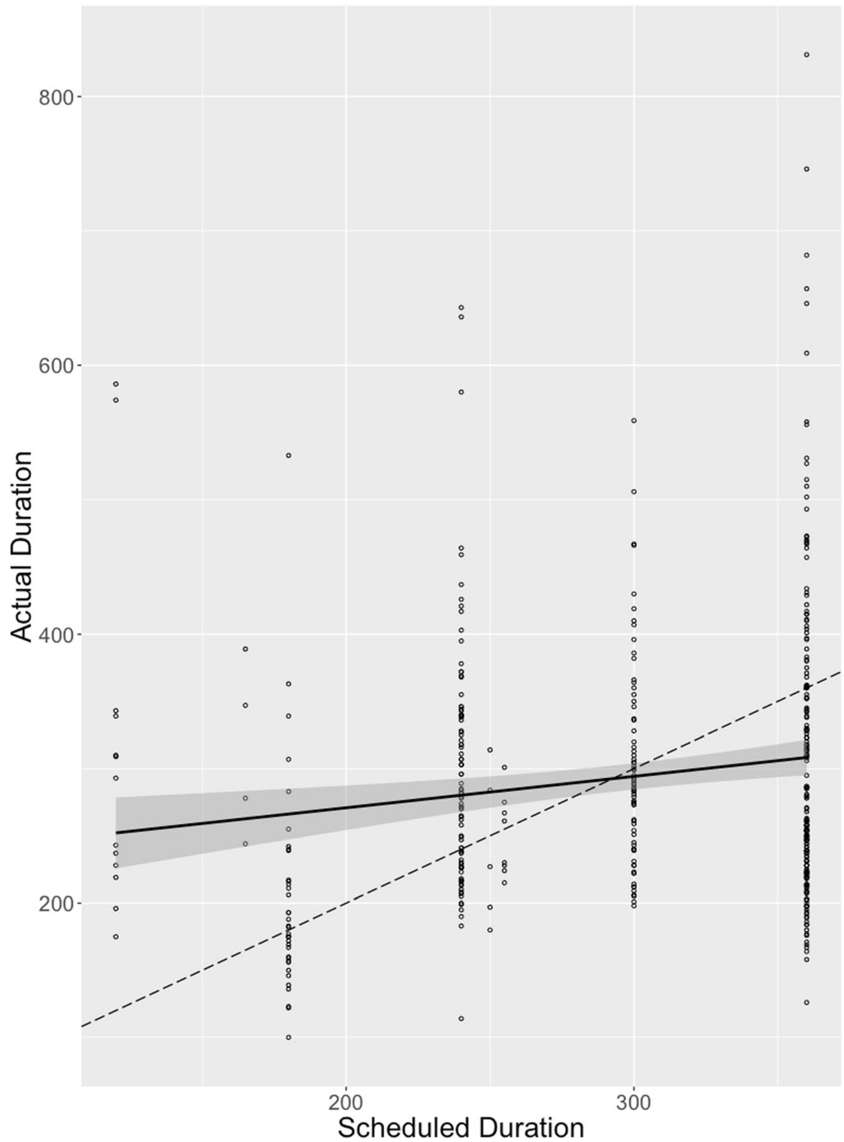

Fig. 1 Scatterplot of the scheduled duration and actual duration with the baseline model (solid line) and the line of perfect prediction (dashed line)

variables were created for categorical variables (with values of 0 or 1), 2) binary variables were converted to value 0 or 1, and 3) continuous variables were normalized by scaling the inputs to have mean 0 and a variance of 1. For the neural network, we developed the model with one hidden layer consisting of 5 nodes. Performance of all models was determined by the average root-mean-squared error (RMSE) calculated from ten-fold cross-validation, followed by bootstrapping ($R = 100$) to determine 95% confidence intervals (CI).

Table 3 Performance of the models with average RMSE from 10-fold cross-validation and bootstrapping

Model	Average RMSE from 10-fold cross-validation	95% Confidence interval
Baseline model	100.4	90.5–110.3
Multivariable linear regression	86.8	78.4–95.1
Ridge regression	82.4	73.3–91.5
Lasso regression	81.3	71.5–91.0
Random forest	81.9	75.8–88.1
Boosted regression tree	80.2	74.0–86.4
Neural network	89.6	71.2–107.9

RMSE root-mean-squared error

In order to test how well our best machine learning algorithm will perform on actual OR cases, we then applied the algorithm to the entire dataset to determine the predicted case length for every case. We then compared the performance of the predicted case length from the machine learning algorithm against the current scheduled durations. One method was to compare the proportion of cases with an accurate booked case length. Currently, case booking accuracy is assessed by whether the actual duration of the case is within 15% of the scheduled duration. For all cases, we determined what

Table 4 Median original scheduled duration, new predicted duration, and actual duration, for all procedure groups

Procedure group	Scheduled duration (mins)	Predicted duration (mins)	Actual duration (mins)
Bowel resection	240	303.6	287
Low anterior resection	360	328.7	328.5
Myotomy	180	167.2	158.5
Cystectomy	360	469.3	466
Partial nephrectomy	240	280.8	278
Radical nephrectomy	240	269.4	293.5
Radical prostatectomy	360	247.2	244
Ureteral reimplantation	240	387.0	395
Trans-oral robotic surgery	120	290.1	293
Salpingo-oophorectomy (benign)	180	230.1	199.5
Simple hysterectomy (benign)	300	291.8	283
TAH-BSO (malignant)	300	299.5	291.5

TAH-BSO Total Abdominal Hysterectomy and Bilateral Salpingo-oophrectomy

proportion of these cases was accurately booked using both the original scheduled duration and the new predicted case length. For both the original scheduled duration and the new predicted duration, we calculated the cumulative minutes that are over- or under-predicted compared to the actual duration by summing the absolute value of the difference. Lastly, we calculated the proportion of cases that were over- or under-predicted for both the original scheduled duration and for the new predicted case lengths. For all analyses, the level of significance was set at $p < 0.05$. Chi-square test was used to compare the number of accurate cases when using the scheduled duration versus the predicted case lengths.

Results

A total of 12 procedural categories and 424 cases were included in the analysis. Table 2 lists the frequency of each covariate among our patient population. After 10-fold cross-validation and bootstrapping, the baseline model for predicting actual case duration yielded a RMSE of 100.4 min (95% CI 90.5–110.3). Figure 1 shows a scatterplot of the actual durations versus the scheduled durations for all cases. The solid line indicates the baseline model, while the dashed line indicates the line of perfect prediction.

MLR improved on the RMSE as compared to the baseline model, with a RMSE of 86.8 min (95% CI 78.4–95.1). The 95% CI of MLR overlaps with the 95% CI of the baseline model, suggesting a non-significant difference. Ridge regression decreased the RMSE to 82.4 min (95% CI 73.3–91.5). Again, this 95% CI overlaps with the baseline model, suggesting a non-significant difference. Lasso regression decreased the RMSE to 81.3 min (95% CI 71.5–91.0), but also overlaps with the baseline model's 95% CI. RF had a slightly higher RMSE of 81.9 min (95% CI 75.8–88.1), but its 95% CI does not overlap with that of the baseline model, indicating that the RF algorithm had a significantly smaller RMSE. BRT decreased the average RMSE to 80.2 min (95% CI 74.0–86.4), indicating slightly better performance as compared to RF. Lastly, NN had a RMSE of 89.6 min (95% CI 71.2–107.9), indicating a non-significant difference from the baseline model. Table 3 summarizes the RMSE of each model.

After determining that the BRT model performed the best, we applied the algorithm to all cases within the cohort to yield a machine-learning generated predicted case length. Table 4 shows the comparison between the median scheduled duration, predicted duration, and actual duration of all procedure groups.

The cumulative minutes over- or under-predicted decreased from 36,894 min using the original scheduled duration to 21,133.86 min using the BRT predictive model

Table 5 Comparison of the cumulative minutes over/under compared to the actual case duration, percent of cases over/under, and booking accuracy rate between the original scheduled duration versus the new predicted duration using boosted regression trees

	Cumulative minutes over/under	% of Cases over/under	Accurately booked cases	*p*-value[a]
Original scheduled duration	36,894	57.5%/42.5%	148 (34.9%)	<0.001
New predicted duration	21,133.86	54.2%/45.8%	219 (51.7%)	

[a] Chi-square test of the booking accuracy rate

(Table 5). The percentage of cases in which the case durations are over- or under-predicted compared to the actual case duration for both the BRT model and the original scheduled duration are also shown in Table 5. Using the BRT predicted case length, the number of accurately booked cases increased from 148 to 219 (34.9% to 51.7%, $p < 0.001$) (Table 5).

Discussion

By using various machine learning approaches, we decreased the RMSE for predicted RAS case lengths at our institution. While all six machine learning models we tested had an average RMSE that was lower than the baseline model, only RF and BRT significantly improved the average RMSE, with BRT performing the best.

Currently, cases at our institution are scheduled by averaging the case lengths of each services' previous 10 cases and by surgeon preference. Our baseline model showed that the current scheduled duration is a poor predictor of actual case duration, resulting in high RMSE. By including 28 variables and applying machine learning techniques, we were able to significantly decrease the RMSE compared to the baseline model. A switch to the BRT model would have decreased the cumulative minutes that were over- or under-predicted by 15,760.14 min. The original scheduled duration was also more likely to over-predict the case duration compared to the BRT model. In general, over-predicting a case duration (in which a case finishes earlier than expected) is more desirable than under-predicting a case duration (in which a case finishes later than expected). However, scheduling a case for longer than it should be decreases overall OR efficiency by increasing wasted block time, staffing, and equipment. When applying the ± 15% buffer that is standard at our institution to gauge booking accuracy, the BRT model significantly increased the booking accuracy by 16.8%. Accurately predicting the case duration is critical in planning and scheduling of cases throughout the day [11]. This is more so for robotic surgery, due to the fact that it is a highly-sought and limited resource.

In the quest to improve OR efficiency, hospitals have looked to team-training, Lean and Six Sigma methodologies, and variability modeling [1, 14–16]. However, it has been suggested that OR efficiency is highest when allocation of OR time is performed at the onset [17]. To that end, advanced statistical modeling and computerized simulations have been tested to assess the impact of different scheduling strategies on OR efficiency [3, 10, 18–21]. However, no studies have used patient factors, disease factors, and operating room factors to construct machine learning models to improve case duration predictions in the scheduling of RAS.

Operationalizing these findings can be potentially challenging. Machine learning is often perceived as an esoteric "black box", where variables are entered into a "machine" and predictions are generated. This misperception may cause stakeholders to greet it with skepticism. However, machine learning has been used in a wide range of medical applications to great effect [22–24], though the clinical application of this technology remains limited in healthcare. In our model, the majority of the variables used to construct the algorithm can be auto-populated from patients' electronic health record (EHR). Therefore, an OR scheduling program utilizing machine learning can be tethered to the EHR and seamlessly schedule cases with minimal (if any) manual data entry (Fig. 2). Though we only included robotic surgeries because it represents a distinct cohort, this technique can be applied to other types of surgery or specialties. Lastly, another advantage of machine learning is the opportunity for the algorithm to "learn" with time, allowing the predictions to become more accurate as the number of samples increases.

Our study has several limitations. In our cross-validation, we randomly divided the cases into 10 blocks to avoid sampling bias. In doing so, cases were no longer sequential in time. Because robotic surgery is an emerging technology, a surgeon's case duration may decrease as he or she becomes more familiar with the platform. We only included robotic surgery in this study because it represents a distinct cohort of cases. To apply this model to more approaches would require a much larger dataset to build an accurate model. In this respect, we were also somewhat limited by the computation power available. Machine learning algorithms can become rather complex, so decisions had to be made to balance model performance with practicality. If we are to operationalize these models, they must perform with accuracy but also in a timely manner. It has also been suggested that some surgeons may artificially make

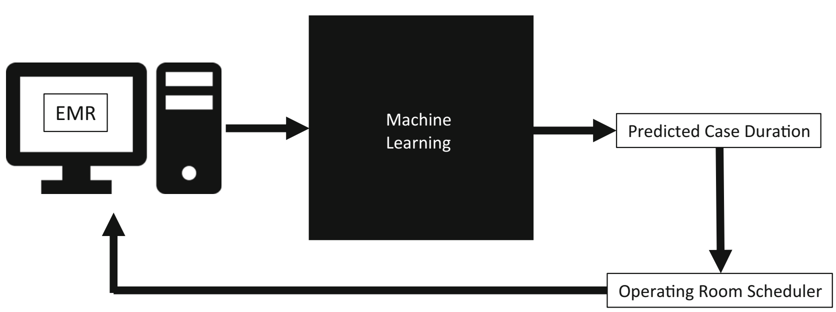

Fig. 2 Schematic diagram of the possible integration of the EHR to a machine learning algorithm for case duration planning

their scheduled durations inaccurate in order to book more cases per day. If this is widespread, then it may explain why the baseline model is so inaccurate. We have no evidence that this occurs frequently at our institution and using our algorithm to schedule cases would take the ability to manipulate the scheduled duration out of the hands of the surgeon. Lastly, the question of generalizability is important to discuss. The models were created using a unique cohort of patients at our institution, so the RMSE findings presented here may not be reproducible at another institution. In addition, because our cross-validation, bootstrapping, and certain mechanisms of some of the machine learning models introduce randomness into the equation, there are small variations in results even when repeating on the same dataset. However, the stochastic nature of our models is a strength of the machine learning, and the methodology employed here is robust enough that these findings should be applicable at other institutions.

In summary, our study demonstrates that predictive modeling using machine learning is more accurate than the current method to predict case duration, and that implementation of a machine learning algorithm can lead to an increase in the number of accurately booked case durations. This improved predictive power may lead to decreases in both over- and under-utilization of operating room resources. Future studies should focus on integrating this technology to the actual scheduling of cases and exploring its effects on OR efficiency and workflow.

Acknowledgements Dr. Beiqun Zhao is funded by the National Library of Medicine Training Grant: NIH grant T15LM011271.

Funding This study was funded by the National Library of Medicine Training Grant (T15LM011271).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

References

1. Cima, R. R., Brown, M. J., Hebl, J. R., Moore, R., Rogers, J. C., Kollengode, A., Amstutz, G. J., Weisbrod, C. A., Narr, B. J., and Deschamps, C., Use of lean and six sigma methodology to improve operating room efficiency in a high-volume tertiary-care academic medical center. *J. Am. Coll. Surg.* 213:83–92, 2011. https://doi.org/10.1016/j.jamcollsurg.2011.02.009.
2. Macario, A., What does one minute of operating room time cost? *J. Clin. Anesth.* 22:233–236, 2010. https://doi.org/10.1016/j.jclinane.2010.02.003.
3. Eijkemans, M. J., van Houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G., Predicting the unpredictable. *Anesthesiology* 112:41–49, 2010. https://doi.org/10.1016/j.jacc.2015.12.063.
4. Intuitive Surgical (2016) Intuitive Surgical, Inc. 2016 Annual Report
5. Higgins, R. M., Frelich, M. J., Bosler, M. E., and Gould, J. C., Cost analysis of robotic versus laparoscopic general surgery procedures. *Surg. Endosc.* 31:185–192, 2017. https://doi.org/10.1007/s00464-016-4954-2.
6. Barbash, G. I., and Glied, S. A., New technology and health care costs - the case of robot-assisted surgery. *N. Engl. J. Med.* 363:701–704, 2010.
7. Zhou, J., Dexter, F., MacArio, A., and Lubarsky, D. A., Relying solely on historical surgical times to estimate accurately future surgical times is unlikely to reduce the average length of time cases finish late. *J. Clin. Anesth.* 11:601–605, 1999. https://doi.org/10.1016/S0952-8180(99)00110-5.
8. Wright, I. H., Kooperberg, C., Bonar, B. A., and Bashein, G., Statistical modeling to predict elective surgery time. *Anesthesiology* 85:1235–1245, 1996. https://doi.org/10.1167/8.5.1.
9. Pandit, J. J., and Carey, A., Estimating the duration of common elective operations: implications for operating list management. *Anaesthesia* 61:768–776, 2006. https://doi.org/10.1111/j.1365-2044.2006.04719.x.
10. Kougias, P., Tiwari, V., and Berger, D. H., Use of simulation to assess a statistically driven surgical scheduling system. *J. Surg. Res.* 201:306–312, 2016. https://doi.org/10.1016/j.jss.2015.10.043.
11. Wu, A., Weaver, M. J., Heng, M. M., and Urman, R. D., Predictive model of surgical time for revision total hip arthroplasty. *J. Arthroplast.* 32:2214–2218, 2017. https://doi.org/10.1016/j.arth.2017.01.056.
12. Okike, K., O'Toole, R. V., Pollak, A. N., Bishop, J. A., McAndrew, C. M., Mehta, S., Cross, W. W., Garrigues, G. E., Harris, M. B., and Lebrun, C. T., Survey finds few orthopedic surgeons know the costs of the devices they implant. *Health Aff.* 33:103–109, 2014. https://doi.org/10.1377/hlthaff.2013.0453.
13. Kayış, E., Wang, H., Patel, M., Gonzalez, T., Jain, S., Ramamurthi, R. J., Santos, C., Singhal, S., Suermondt, J., and Sylvester, K., Improving prediction of surgery duration using operational and temporal factors. *AMIA Ann. Symp. Proc.*:456–462, 2012.
14. Smith, C. D., Spackman, T., Brommer, K., Stewart, M. W., Vizzini, M., Frye, J., and Rupp, W. C., Re-engineering the operating room using variability methodology to improve health care value. *J. Am. Coll. Surg.* 216:559–570, 2013. https://doi.org/10.1016/j.jamcollsurg.2012.12.046.
15. Tagge, E. P., Thirumoorthi, A. S., Lenart, J., Garberoglio, C., and Mitchell, K. W., Improving operating room efficiency in academic children's hospital using lean six sigma methodology. *J. Pediatr. Surg.* 52:1040–1044, 2017. https://doi.org/10.1016/j.jpedsurg.2017.03.035.
16. Strum, D. P., Sampson, A. R., May, J. H., and Vargas, L. G., Surgeon and type of anesthesia predict variability in surgical procedure times. *Anesthesiology* 92:1454–1466, 2000. https://doi.org/10.1097/00000542-200005000-00036.
17. Dexter, F., and Epstein, R. H., Operating room efficiency and scheduling. *Curr. Opin. Anaesthesiol.* 18:195–198, 2005. https://doi.org/10.1097/01.aco.0000162840.02087.15.
18. Stepaniak, P. S., Heij, C., Mannaerts, G. H. H., De Quelerij, M., and De Vries, G., Modeling procedure and surgical times for current procedural terminology-anesthesia-surgeon combinations and evaluation in terms of case-duration prediction and operating room efficiency: a multicenter study. *Anesth. Analg.* 109:1232–1245, 2009. https://doi.org/10.1213/ANE.0b013e3181b5de07.
19. Dexter, F., Macario, A., Traub, R. D., Hopwood, M., and Lubarsky, D. A., An operating room scheduling strategy to maximize the use of operating room block time: computer simulation of patient scheduling and survey of patients' preferences for surgical waiting time. *Anesth. Analg.* 89:7–20, 1999. https://doi.org/10.1097/00000539-199907000-00003.

20. Tiwari, V., Dexter, F., Rothman, B. S., Ehrenfeld, J. M., and Epstein, R. H., Explanation for the near-constant mean time remaining in surgical cases exceeding their estimated duration, necessary for appropriate display on electronic white boards. *Anesth. Analg.* 117:487–493, 2013. https://doi.org/10.1213/ANE.0b013e31829772e9.
21. Dexter, F., Ledolter, J., Tiwari, V., and Epstein, R. H., Value of a scheduled duration quantified in terms of equivalent numbers of historical cases. *Anesth. Analg.* 117:205–210, 2013. https://doi.org/10.1213/ANE.0b013e318291d388.
22. Bejnordi, B. E., Veta, M., Van Diest, P. J., Van Ginneken, B., Karssemeijer, N., Litjens, G., Van Der Laak, J. A. W. M., Hermsen, M., Manson, Q. F., Balkenhol, M., Geessink, O., Stathonikos, N., Van Dijk, M. C. R. F., Bult, P., Beca, F., Beck, A. H., Wang, D., Khosla, A., Gargeya, R., Irshad, H., Zhong, A., Dou, Q., Li, Q., Chen, H., Lin, H. J., Heng, P. A., Haß, C., Bruni, E., Wong, Q., Halici, U., Öner, M. Ü., Cetin-Atalay, R., Berseth, M., Khvatkov, V., Vylegzhanin, A., Kraus, O., Shaban, M., Rajpoot, N., Awan, R., Sirinukunwattana, K., Qaiser, T., Tsang, Y. W., Tellez, D., Annuscheit, J., Hufnagl, P., Valkonen, M., Kartasalo, K., Latonen, L., Ruusuvuori, P., Liimatainen, K., Albarqouni, S., Mungal, B., George, A., Demirci, S., Navab, N., Watanabe, S., Seno, S., Takenaka, Y., Matsuda, H., Phoulady, H. A., Kovalev, V., Kalinovsky, A., Liauchuk, V., Bueno, G., Fernandez-Carrobles, M. M., Serrano, I., Deniz, O., Racoceanu, D., and Venâncio, R., Diagnostic assessment of deep learning algorithms for detection of lymph node metastases in women with breast cancer. *JAMA* 318:2199–2210, 2017. https://doi.org/10.1001/jama.2017.14585.
23. Upstill-Goddard, R., Eccles, D., Fliege, J., and Collins, A., Machine learning approaches for the discovery of gene-gene interactions in disease data. *Brief. Bioinform.* 14:251–260, 2013. https://doi.org/10.1093/bib/bbs024.
24. Litjens, G., Kooi, T., Bejnordi, B. E., Setio, A. A. A., Ciompi, F., Ghafoorian, M., van der Laak, J. A. W. M., van Ginneken, B., and Sánchez, C. I., A survey on deep learning in medical image analysis. *Med. Image Anal.* 42:60–88, 2017. https://doi.org/10.1016/j.media.2017.07.005.