



Machine Learning Can Improve Estimation of Surgical Case Duration: A Pilot Study

Justin P. Tuwatananurak¹ · Shayan Zadeh² · Xinling Xu¹ · Joshua A. Vacanti¹ · William R. Fulton³ · Jesse M. Ehrenfeld⁴ · Richard D. Urman^{1,5} 

Received: 24 June 2018 / Accepted: 8 January 2019 / Published online: 17 January 2019
© Springer Science+Business Media, LLC, part of Springer Nature 2019

Abstract

Operating room (OR) utilization is a significant determinant of hospital profitability. One aspect of this is surgical scheduling, which depends on accurate predictions of case duration. This has been done historically by either the surgeon based on personal experience, or by an electronic health record (EHR) based on averaged historical means for case duration. Here, we compare the predicted case duration (pCD) accuracy of a novel machine-learning algorithm over a 3-month period. A proprietary machine learning algorithm was applied utilizing operating room factors such as patient demographic data, pre-surgical milestones, and hospital logistics and compared to that of a conventional EHR. Actual case duration and pCD (Leap Rail vs EHR) was obtained at one institution over the span of 3 months. Actual case duration was defined as time between patient entry into an OR and time of exit. pCD was defined as case time allotted by either Leap Rail or EHR. Cases where Leap Rail was unable to generate a pCD were excluded. A total of 1059 surgical cases were performed during the study period, with 990 cases being eligible for the study. Over all sub-specialties, Leap Rail showed a 7 min improvement in absolute difference between pCD and actual case duration when compared to conventional EHR ($p < 0.0001$). In aggregate, the Leap Rail method resulted in a 70% reduction in overall scheduling inaccuracy. Machine-learning algorithms are a promising method of increasing pCD accuracy and represent one means of improving OR planning and efficiency.

Keywords Operating room · Case duration · Prediction · Model · Natural language processing · Efficiency

Introduction

Up to 60% of hospitalized patients will eventually require surgical intervention, making the operating room (OR) simultaneously a significant source of revenue and overhead [1, 2].

One factor that contributes to a significant overhead cost is the inefficient use of OR time [3]. OR utilization is a metric that is often used to determine the efficiency of OR use and compared against known benchmarks [4]. Inadequate estimation of case duration can lead to both under- and overutilization of the OR time.

The process of surgical scheduling begins by compiling a list of cases and their predicted durations. If cases consistently run longer than anticipated, OR overutilization will result in costly overtime pay and staff dissatisfaction. Even worse, if case times are consistently shorter than expected, OR underutilization will result in increased staff idle time, which is associated with up to a 60% higher cost [5]. Furthermore, scheduling inefficiencies often have downstream effects on various performance metrics (i.e., length of hospital stay, patient satisfaction), which can have large ramifications on patient outcomes and hospital reimbursement.

One common approach to predicting case duration places the responsibility upon the surgeon, who personally reserves a block of OR time based on surgical approach, patient

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

✉ Richard D. Urman
rurman@bwh.harvard.edu

¹ Department of Anesthesiology, Perioperative and Pain Medicine, Brigham and Women's Hospital, Boston, MA 02115, USA

² Leap Rail Inc, Houston, TX, USA

³ NorthBay Healthcare, Fairfield, CA, USA

⁴ Department of Anesthesiology, Vanderbilt University School of Medicine, Nashville, TN, USA

⁵ Center for Perioperative Research, Brigham and Women's Hospital, Boston, MA, USA

comorbidities, and clinician expertise. With this method, surgeons overestimate case duration up to 32% of the time, and underestimate 42% of the time [6]. Another common approach uses the electronic health record (EHR) to calculate a case duration based on historical data for a given procedure and/or surgeon. When comparing these two approaches, EHR-generated case times have been shown to have a modestly higher accuracy [7]. However, commercially available EHRs only generate case durations for the average patient – it does not take patient factors (e.g. age, sex, body mass index, allergies, ASA Physical Status classification and associated comorbidities), procedure-specific considerations (e.g. implant type, use of invasive monitoring, anesthesia type), hospital logistics (staff, equipment, time of day, day of the week), or prior milestones (e.g. case delays, cancellations, turnover times) into account, despite studies showing that these factors can have up to a 30% impact on total case duration [8, 9].

A third approach that is rarely employed is to combine the first two approaches by combining clinician input with EHR-generated predicted case duration (pCD), defined as the predicted time between a patient's entry into the OR and their exit. A more novel approach to this would be to utilize machine-learning with natural language processing to leverage pre-existing data in the EHR in order to tailor pCD to each patient. Our study compares the performance of one such novel algorithm to the conventional methods of using historical means. We hypothesized that there will be a significant improvement in pCD accuracy. If this is the case, this new approach to predicting case duration may represent an additional means of improving OR efficiency and facilitating scheduling of cases.

Methods

Feature selection

Formation of the algorithm began with feature selection. To do so, data at the institutional level were pushed through a series of Extraction, Transformation, and Loading processes (ETL) in order to correlate and optimize data for consumption by the machine-learning engine. The original dataset going through this ETL process for each case included patient information, providers involved in the case, facility details, procedure being performed, as well as prior events in the surgical suite. Table 1 summarizes some of the information processed in this step. All of the data points that are input into the model came from existing documentation about surgical cases in an EHR system. Given that these data can be in different formats across different EHR systems (or even different versions of the same EHR system), the Leap Rail platform is designed such that it is capable of extracting this information from a variety of different input formats and systems. These raw data

are then staged and transformed into a tab-separated values file format and fed into the machine learning model. The output is a simple numeric prediction for the length of surgical cases which can be published in a variety of formats, depending on the needs of the consuming system.

Appropriate numeric representations were assigned to non-numeric informative attributes such as free text surgeon comments in the EHR. This information was analyzed in order to build a corresponding set of relevant individual measurable properties: *features* [10]. Choosing informative, discriminating and independent features is a crucial step for an effective algorithm. This is an iterative process and the resulting features can vary by the dataset being used for training. For the machine-learning algorithm used in this study, over 1500 features were identified and subsequently fed into various machine-learning algorithms such as gradient-boosted tree regression, decision trees, and random forests [11–13].

Model development and training

The proprietary Leap Rail® engine uses a combination of supervised learning algorithms. A supervised learning algorithm makes predictions based on a set of examples [14]. Each example used for training is labeled with the value of interest. For this study, the examples given to the supervised learning algorithms were historical surgical cases performed at the institution (n ~15,000), and the value of interest was the actual case duration.

A supervised learning algorithm looks for patterns in those labels. It can use any information that might be relevant among all the provided features in training data. Each algorithm looks for different types of patterns. After the algorithm finds the best patterns it can, it uses that pattern in making predictions for future cases.

In order to select the best algorithm for the cases performed at the institution, historical case data were split into training and test sets. The Leap Rail engine was exposed to the training data in order to create multiple models using different algorithms and feature sets, and their performance was subsequently measured against unseen surgical cases using the test set. Model performance was objectively compared, and the most accurate model was selected for use at the institution.

The winning model was deployed and used for predicting duration of unseen future surgical cases. In this study, we have evaluated the accuracy of pCD generated by the Leap Rail engine on these unseen cases compared to the EHR pCD predictions.

Study data/variables collected

All data were received from Leap Rail, Inc. (San Francisco, CA, USA). All potential patient or clinician identifiers were removed prior to data transmission and analysis. For this

Table 1 Examples of OR case-related factors (patient, providers, facility/room, procedure, prior events) included in the machine-learning algorithm

| Patient | Providers | Facility / Room | Procedure | Prior Events |
|------------------------------------|---|---------------------|-------------------------|--|
| Age, sex | Surgeon(s) | Hospital bed census | Procedure Type | Last food/drink intake |
| Height, weight | Anesthesiologist(s) | Equipment | Surgeon comments | Timing of prior perioperative milestones |
| Allergies | Scrub nurse(s) and technicians(s) | Staff | Procedure modifiers | Case delays |
| Medical conditions | Circulator nurse(s) and technicians(s) | Day of the week | Anesthesia type | Cancellations |
| ASA physical status classification | Whether the assigned group has worked on this type of case before | Time of day | Implants / tissues used | Room turnover time |

OR = operating room; ASA = American Society of Anesthesiologists

study, we examined all operative cases performed at NorthBay Healthcare, a Level II Trauma center located in Solano County, California from January to March 2018. Raw data were collected and subsequently provided by Leap Rail, Inc., the developer of the machine-learning algorithm used for this study. Variables collected from each case included primary procedure, subspecialty, EHR-predicted duration, and actual case duration. EHR-predicted case times were generated using the Cerner System (North Kansas City, MO, USA) currently used at the study facility, which did so by using historical means. Actual case duration was defined as the patient in-room time to patient out-of-room time. Each procedure name was determined by its corresponding ICD-10 code, while the primary specialty for each case was determined by recording the booking surgeon’s credentials. Performing surgeons were aware of Leap Rail pCD as well as the pCD generated by the existing EHR (Cerner).

Surgical cases were excluded from the study if the Leap Rail engine was unable to generate a predicted case time due to insufficient training data. There were no further exclusion criteria for this study. Microsoft Excel 2016 (Redmond, WA, USA) was used for table and figure production.

Statistical methods

The primary outcome was the absolute difference between the actual duration and the predicted value (i.e. prediction error). It was summarized using median value and interquartile range (Q1-Q3), and was compared using Wilcoxon rank-sum test between Leap Rail and EHR groups. Additionally, for each subspecialty, the prediction error was compared using the same test and unadjusted *p* values were reported. When defining having prediction error ≤ 15 min as acceptable, a Chi-square test was performed to compare whether there was a significant difference of the number of acceptable predictions in each group. All analyses were performed using R software version 3.4.1. Statistical tests were two-sided with α = 0.05.

Results

A total of 1059 surgical cases were performed at NorthBay Healthcare during the study period of three months. Leap Rail was able to generate a predicted time for 990 of these cases, which represented approximately 93.5% of total surgical volume. No further cases were excluded. The most heavily represented subspecialties were Gastroenterology (*n* = 207, 20.9%), General Surgery (*n* = 240, 24.2%), and Orthopedics (*n* = 166, 16.8%). The least represented were Pulmonary (*n* = 1, 0.1%), Oral/Maxillofacial (*n* = 2, 0.2%), and Interventional Radiology (*n* = 4, 0.4%).

For all cases, the median absolute difference between actual and predicted case times was 20 min for Leap Rail (Q1-Q3, 9–40) and 27 min (Q1-Q3, 13–50) for EHR (Cerner) (Table 2), demonstrating a 7-min improvement for Leap Rail predictions (*p* value < 0.0001). A density distribution plot (Fig. 1) comparing the frequency of pCD error by group further illustrates this improvement. Here, the curve for Leap Rail predictions is shifted towards the left, indicating both an increased likelihood of smaller error and a decreased likelihood of larger error when compared to the EHR. Fig. 2 and Fig. 3 graph these data in a scatterplot, with the straight line representing a theoretical perfect relationship between predicted and actual case durations. Here, Leap Rail had a higher Pearson correlation compared to the EHR, again indicating increased accuracy. There is also seen the tendency of EHR pCD to fall within the same time groups despite their actual durations, indicating a level of inflexibility in generated case times.

When broken down by subspecialty, Leap Rail was more accurate for 14 of the 16 subspecialties; however, of those 14 subspecialties, only the findings for Gastroenterology, General Surgery, Orthopedics, and Urology were statistically significant (Table 2). Subsequently, Fig. 4 shows the box plot of prediction error for these four specialties as well as all cases. Here, the median absolute differences are consistently higher in the EHR group when compared to the Leap Rail group. However, both groups still had significant amount of outliers, indicating intra-operative variability that was difficult to account for regardless of prediction modality.

Table 2 Median value of absolute difference, with interquartile ranges, between predicted values and actual values by groups

| Service | # of cases | EHR: Absolute Median (Q1-Q3) Difference | Leap Rail: Absolute Median (Q1-Q3) Difference | P value |
|----------------------------|------------|---|---|---------|
| Cardiovascular | 19 | 79.0 (31.0–143.5) | 28.0 (7.0–111.0) | 0.0558 |
| Otorhinolaryngology | 11 | 37.0 (28.0–50.5) | 26.0 (15.5–48.0) | 0.4299 |
| Gastroenterology | 207 | 22.0 (11.0–35.0) | 17.0 (7.5–28.0) | 0.0039 |
| General Surgery | 240 | 25.0 (13.0–47.5) | 17.0 (8.5–35.5) | 0.0002 |
| Gynecology | 65 | 29.0 (14.0–55.0) | 24.0 (11.0–50.0) | 0.1582 |
| Interventional Radiology | 4 | 22.5 (15.8–28.8) | 7.5 (4.8–11.5) | 0.1143 |
| Neurosurgery | 19 | 14 | 24 | 0.7812 |
| Oral/Maxillofacial | 2 | 23.0 (20.0–26.0) | 23.5 (22.8–24.3) | 1 |
| Orthopedics | 166 | 35.0 (14.0–74.0) | 25.0 (13.0–48.0) | 0.0088 |
| Plastic Surgery | 37 | 28.0 (13.0–59.0) | 26.0 (9.0–55.5) | 0.9010 |
| Podiatry | 51 | 18.0 (6.5–36.5) | 15.0 (8.0–27.0) | 0.7126 |
| Interventional Pulmonology | 1 | 25 | 5 | 1 |
| Spine Surgery | 29 | 58.0 (11.0–87.0) | 43.0 (22.0–76.0) | 0.9380 |
| Thoracic Surgery | 6 | 72.0 (36.8–75.8) | 34.0 (18.8–51.5) | 0.1488 |
| Urologic Surgery | 85 | 27.0 (14.0–44.0) | 12.0 (7.0–33.0) | 0.0022 |
| Vascular Surgery | 48 | 27.0 (17.0–44.5) | 26.5 (14.8–41.3) | 0.7195 |
| All | 990 | 27.0 (13.0–50.0) | 20.0 (9.0–39.8) | <0.0001 |

EHR = electronic health record; Q3 - Q1 = Interquartile range

For a single case pCD, predictions were 31.2% and 41.1% accurate for EHR and Leap Rail, respectively, when defining the threshold for a clinically significant prediction error for a single case as 15 min ($p < 0.0001$) (Table 3). When looking at case duration in aggregate over a 3-month period, the 990 observed cases amounted to 110,130 min of operating room time. EHR-predicted total surgery time was 79,435 min compared to the Leap Rail predicted time of 98,554 min. Overall,

this resulted in a 70% reduction in scheduling inaccuracy for the operating suite (Table 4).

Discussion

In this study, we compared the accuracy of Leap Rail surgical case time prediction to that of conventional EHR approach. When examining all cases during the study period, our results showed a statistically significant improvement of approximately 7 min per case. When broken down by subspecialty, the most prominent improvements were seen in cardiology (although this effect was not significant), followed by orthopedics and urology, which had up to a 15-min improvement in accuracy. Conventional EHR outperformed Leap Rail for two subspecialties (neurology and maxillofacial), but these differences were not statistically significant.

There is a difference between statistical and clinical significance that needs to be addressed. In fact, one may be skeptical about the clinical significance of a 7-min improvement in accuracy per case. However, even when setting a threshold of 15 min as a clinically significant error (a more typical unit of time used in OR scheduling), Leap Rail still represented a 10% improvement in accuracy over conventional means. Furthermore, as some ORs can have up to 6–7 cases per day, the cumulative effect may ultimately represent a significant financial advantage if it is possible to schedule an additional case [15]. This is best reflected by looking at the compounding effect of total OR cases through the 3-month

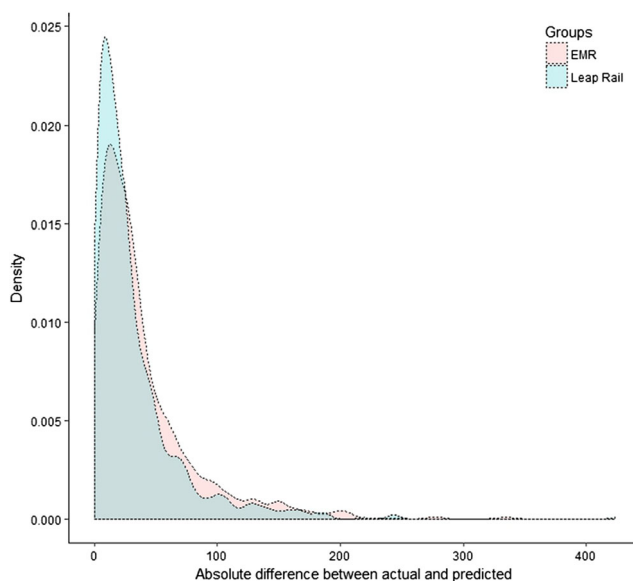
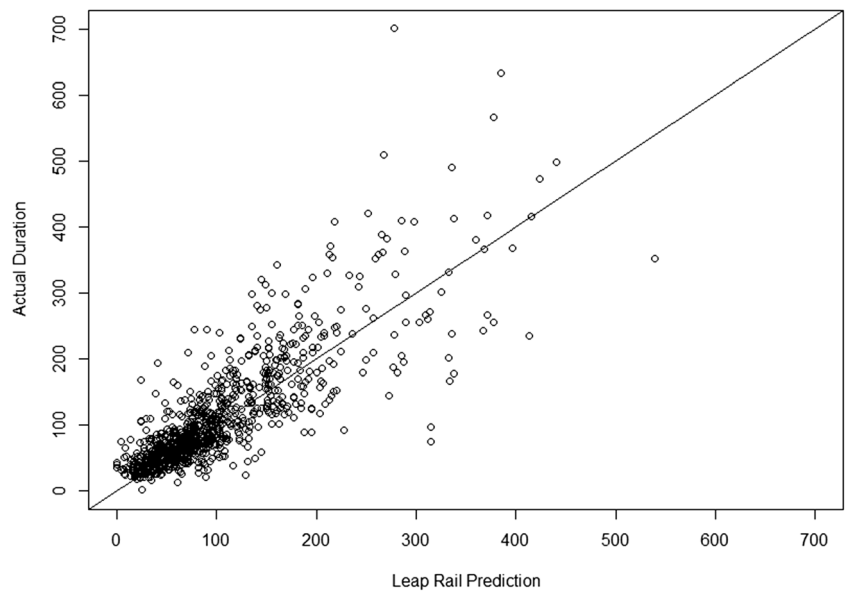


Fig. 1 Density plot of the absolute difference of actual values and predictions by groups

Fig. 2 Scatterplot of actual duration versus Leap Rail prediction, with Pearson correlation and the correlation = 1 line added



study period: Leap Rail represented a ~19,000 min, or a 70% reduction in scheduling inaccuracy overall.

Previous studies have estimated the cost per OR minute to the hospital to be between \$22 and \$133 per minute [16]. Therefore, by appropriately utilizing the ORs, this would result in a significant reduction in costs over the course of 3 months of this pilot study. This estimate represents a direct savings through OR staffing costs, and actually likely underestimates the cumulative effect seen downstream throughout the hospital system, which includes decreased length of hospital stay due to delays/cancellations, etc. Therefore, despite a seemingly insignificant time savings per case, the cumulative effect could certainly lead to a significant increase in revenue.

While this study focuses on total OR case duration, it is notable that Leap Rail is able to sub-categorize each case into

different surgical, anesthesia, and turnover times. This ability to account for variance within each sub-segment is one of Leap Rail’s biggest advantages over current scheduling modalities. However, we did not directly assess Leap Rail’s performance in this capacity, as the EHR did not generate discrete values for each of these sub-segments. Furthermore, it is ultimately total case duration that is the main determinant of surgical scheduling. Nonetheless, we expect this capability to be reflected in this study via the improvements seen with Leap Rail’s pCD accuracy.

With this in mind, we believe that future studies are warranted to further validate our results. In particular, it would be reasonable to apply the Leap Rail algorithm more broadly among several institutions, and for a longer period of time. The benefits of this would be twofold – one could expect to see more

Fig. 3 Scatterplot of actual duration versus EHR prediction, with Pearson correlation and the correlation = 1 line added

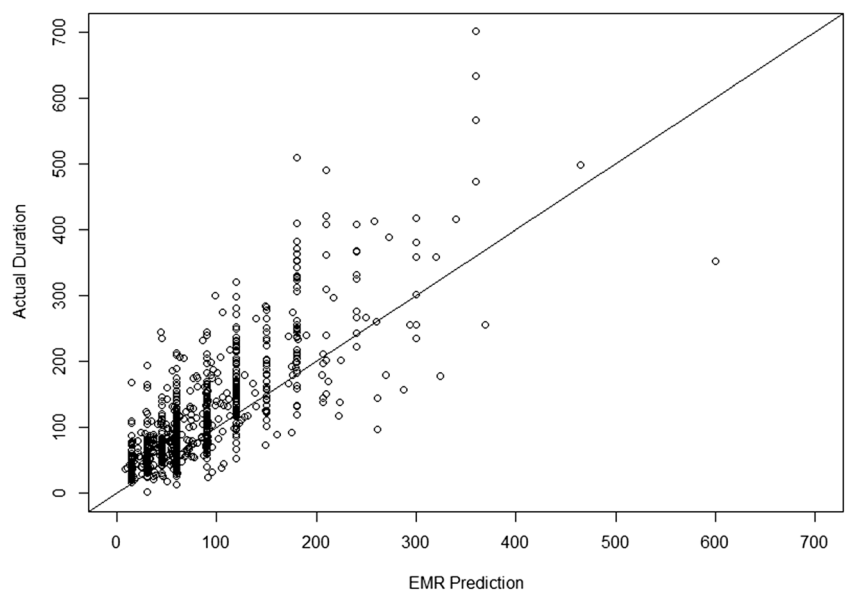
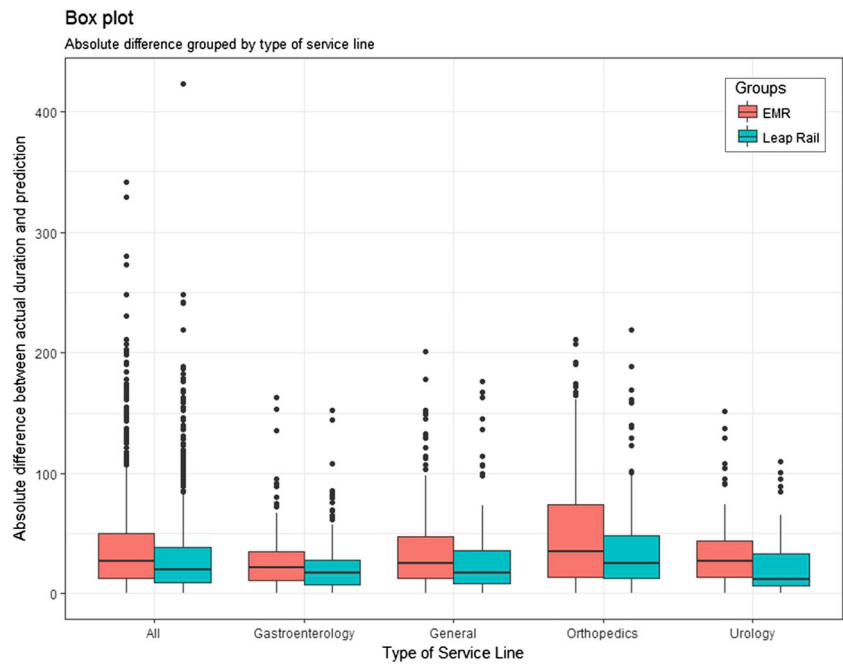


Fig. 4 Box plot of the absolute differences by subspecialty and all cases



generalizability as Leap Rail is exposed to a wider array of cases, and one could also expect to see more accurate predictions as additional training data are incorporated into the machine learning process. Furthermore, future studies could further investigate the accuracy of Leap Rail predictions for anesthesia and turnover time. By doing so, one may be able to determine additional avenues in which OR inefficiency can be minimized.

There have been several prior studies investigating potential means of optimizing surgical scheduling. Many of these studies have established the use of historical means as a superior method of predicting surgical case duration [7, 8]; other studies have investigated using mathematical modeling to improve case sequencing [17, 18] or predicting cases durations when having sparse historical data [19]. However, none of these studies utilized a similar multi-modal approach towards predicting surgical case durations. Furthermore, none of these models were built utilizing a machine-learning algorithm that would be expected to improve with the incorporation of additional training data.

Table 3 A 2-by-2 table demonstrating when prediction error ≤ 15 min was defined as acceptable

| | | Leap Rail Prediction Error > 15 min | | Total |
|---------------------------|-----|-------------------------------------|-------------|-------------|
| | | Yes | No | |
| EHR Prediction Error > 15 | Yes | 472 | 209 | 681 (68.8) |
| | No | 111 | 198 | 309 (31.2)* |
| Total | | 583 (58.9) | 407 (41.1)* | 990 |

* $p < 0.0001$; EHR = electronic health record

There are some limitations to our methodology. A potential weakness of the Leap Rail engine is that it relies on a training set in order to generate a predicted case time. Therefore, it may be unable to do so for procedures that are rare or difficult to classify (multi-specialty surgeries, or combined procedures). Nonetheless, these types of procedures are the minority within any hospital system and are unlikely to represent a significant proportion of OR utilization. Furthermore, these weaknesses are also applicable to conventional EHR case time prediction, which also relies on historical case data. This is best reflected in the 93.5% total utilization rate of Leap Rail, which shows that it is largely applicable to a significant portion of surgical volume.

Another weakness of the designed study was a relatively small sample size. While there was statistically significant improvement in many subspecialties, there were others where the benefit of using Leap Rail was less validated. This was particularly magnified in specialties such as Pulmonary, Oral/Maxillofacial, or Interventional Radiology, which had very few available cases for comparison. However, this study was not intended to be an exhaustive comparison between EHR and Leap Rail generated surgical case times; rather, this

Table 4 Cumulative differences between predicted vs actual case times (in minutes)

| | Actual | EHR | Leap Rail |
|--------------------|---------|--------|-----------|
| Total Minutes | 110,130 | 79,435 | 98,554 |
| Inaccurate Minutes | | 30,695 | 11,576 |
| Inaccuracy Percent | | 38.6% | 11.7% |

EHR = electronic health record

study was designed as a “proof of concept” to demonstrate the potential value of taking a more multimodal approach to case time prediction.

Machine learning models are by definition biased towards their training data which in the case of the model in this study includes information about patients, procedures, healthcare providers, and the facility itself. This model is designed to be able to make high accuracy predictions for many new patients. Leap Rail regularly re-trains their models at each facility with the addition of new case data, so that the new models can account for new team members or even changes in physical topology of the facility or clinical workflows. Also, the model reviewed in this study was custom made for a given organization using training data about its patients, procedures, physicians, nurses, and the facility itself. Therefore, the same exact model is not meant to, and will not perform well if simply applied to a different organization. However, the same methodology and the “know-how” can be applied to training data from other organizations with the expectation of achieving similar results.

In summary, our study compared the accuracy of pCDs generated by an EHR with that of a novel machine-learning algorithm. Our results were statistically significant, and indicated that the algorithm had an average improvement of 7 min per case. This represented an aggregate 70% reduction in scheduling inaccuracy over a three-month period. In trialing this approach, we found that the Leap Rail engine represents a significant improvement in reducing prediction error for surgical case duration. Application of such an algorithm could potentially lead to significant savings for the hospital.

Funding None. Data were provided by Leap Rail, Inc.

Compliance with Ethical Standards

Conflict of Interest Justin P. Tuwatananurak declares that he has no conflict of interest. Shayan Zadeh serves as the CEO of Leap Rail, Inc. Xinling Xu PhD declares that she has no conflict of interest. Joshua A. Vacanti declares that he has no conflict of interest. William R. Fulton declares that he has no conflict of interest. Jesse M. Ehrenfeld declares that he has no conflict of interest. Richard D. Urman has received funding from Medtronic, Merck and Mallinckrodt for unrelated research and honorarium from 3 M and Posimir.

Ethical Approval All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and with the 1964 Helsinki declaration and its later amendments or comparable ethical standards.

Informed Consent Informed consent requirement was waived given the deidentified nature of the data and observational nature of the study.

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

References

- Gordon, T., Paul, S., Lyles, A., and Fountain, J., Surgical unit time utilization review: Resource utilization and management implications. *J. Med. Syst.* 12(3):169–179, 1988.
- Peltokorpi, A., How do strategic decisions and operative practices affect operating room productivity? *Health Care Manag. Sci.* 14(4): 370–382, 2011.
- Gabriel, R. A., Wu, A., Huang, C. C., Dutton, R. P., and Urman, R. D., National incidences and predictors of inefficiencies in perioperative care. *J. Clin. Anesth.* 31:238–246, 2016.
- May, J. H., Spangler, W. E., Strum, D. P., and Vargas, L. G., The surgical scheduling problem: Current research and future opportunities. *Prod. Oper. Manag.* 20(3):392–405, 2011.
- Tankard, K., Acciavatti, T. D., Vacanti, J. C. et al., Contributors to operating room underutilization and implications for hospital administrators. *Health Care Manag. (Frederick).* 37(2):118–128, 2018.
- Laskin, D. M., Abubaker, A. O., and Strauss, R. A., Accuracy of predicting the duration of a surgical operation. *J. Oral Maxillofac. Surg.* 71(2):446–447, 2013.
- Wu, A., Huang, C. C., Weaver, M. J., and Urman, R. D., Use of historical surgical times to predict duration of primary Total knee arthroplasty. *J. Arthroplasty.* 31(12):2768–2772, 2016.
- Stepaniak, P. S., Heij, C., Mannaerts, G. H., De quelerij, M., and DeVries, 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(4):1232–1245, 2009.
- Eijkemans, M. J., Van houdenhoven, M., Nguyen, T., Boersma, E., Steyerberg, E. W., and Kazemier, G., Predicting the unpredictable: A new prediction model for operating room times using individual characteristics and the surgeon's estimate. *Anesthesiology.* 112(1): 41–49, 2010.
- Bishop, C., Pattern recognition and machine learning. Berlin: Springer, 2006, ISBN 0-387-31073-8.
- Mason, L., Baxter, J. Bartlett, P. L., and Freaun, M., Boosting algorithms as gradient descent. In S.A. Solla and T.K. Leen and K. Müller. Advances in neural information processing systems 12. MIT Press. 512–518, 1999.
- Rokach, Lior, and Maimon, O., Data mining with decision trees: Theory and applications. World Scientific Pub Co Inc. ISBN 978-9812771711, 2008.
- Breiman, L. Machine Learning 45: 5, 2001. <https://doi.org/10.1023/A:1010933404324>.
- Russell, S. J., and Norvig, P., Artificial intelligence: A modern approach. Pearson, 2016.
- Dexter, F., and Macario, A., Decrease in case duration required to complete an additional case during regularly scheduled hours in an operating room suite: A computer simulation study. *Anesth. Analg.* 88(1):72–76, 1999.
- Macario, A., What does one minute of operating room time cost? *J. Clin. Anesth.* 22(4):233–236, 2010.
- Van Houdenhoven, M., Van oostrum, J. M., Hans, E. W., Wullink, G., and Kazemier, G., Improving operating room efficiency by applying bin-packing and portfolio techniques to surgical case scheduling. *Anesth. Analg.* 105(3):707–714, 2007.
- Dexter, F., and Traub, R. D., How to schedule elective surgical cases into specific operating rooms to maximize the efficiency of use of operating room time. *Anesth. Analg.* 94(4):933–942, 2002 table of contents.
- Dexter, F., and Ledolter, J., Bayesian prediction bounds and comparisons of operating room times even for procedures with few or no historic data. *Anesthesiology.* 103(6):1259–1167, 2005.