

## Supplementary Info

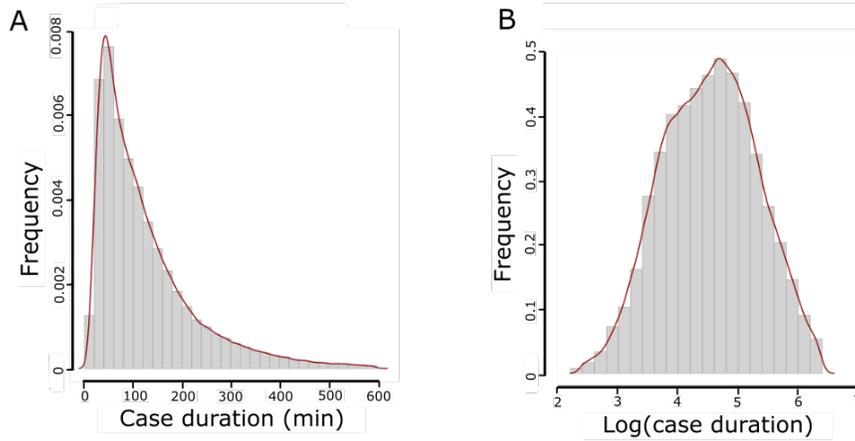


Figure S1: Log transformation of case duration converted the distribution of surgical case duration from (A) skewing to the right to (B) a more normal distribution.

Model	Training set					Internal Testing set					External Testing set				
	$R^2$ (%)	MAE	U (%)	O (%)	W (%)	$R^2$ (%)	MAE	U (%)	O (%)	W (%)	$R^2$ (%)	MAE	U (%)	O (%)	W (%)
Surgeon-specific	31	60.4	50	31	19	30	60.3	50	31	19	30	64.6	52	32	17
Procedure-specific	68	37.3	37	25	37	66	38	38	25	36	66	40.7	40	25	35
Reg	80	30.3	32	25	43	78	31.1	32	25	42	79	33	36	23	41
logReg	84	28.7	24	27	49	83	29.5	25	27	48	84	31	27	25	48
RF	83	29.2	23	27	49	83	28.9	23	27	50	84	30.9	26	26	48
XGB	85	27.3	23	26	51	84	28.7	23	27	49	85	30.2	27	25	48

Table S1: Performance of all the models in the training and internal and external testing sets. The models that were included for comparison in this study were average models for surgeon- or procedure-specific scenarios, multivariate linear regression (Reg), log transformed multivariate linear regression (logReg), random forest (RF) and extreme gradient boosting (XGB). MAE: mean absolute error; U: underage; O: overage; W: within

Predictor variables	Correlation coefficient	Standard error	t-value	p-value
No. of previous urgent and emergent surgeries performed by the same surgeon on the same day	$1.49 \times 10^{-2}$	$6.71 \times 10^{-3}$	2.223	0.03
No. of previous surgeries performed by the surgeon on the same day	$-8.09 \times 10^{-3}$	$7.34 \times 10^{-4}$	-11.029	$<2 \times 10^{-16}$
Total surgical minutes performed by the surgeon on the same day	$6.06 \times 10^{-5}$	$6.10 \times 10^{-6}$	9.946	$<2 \times 10^{-16}$
No. of previous surgeries performed by the surgeon within the last 7 days	$-1.82 \times 10^{-3}$	$3.59 \times 10^{-4}$	-5.08	$3.78 \times 10^{-7}$
Total surgical minutes performed by the surgeon within the last 7 days	$1.65 \times 10^{-5}$	$2.84 \times 10^{-6}$	5.81	$6.22 \times 10^{-9}$

Table S2: Correlation coefficient, standard error, t-value and p-value of predictor variables extracted from primary surgeons' data. This information was obtained from the log-transformed multivariate regression (logReg) model.

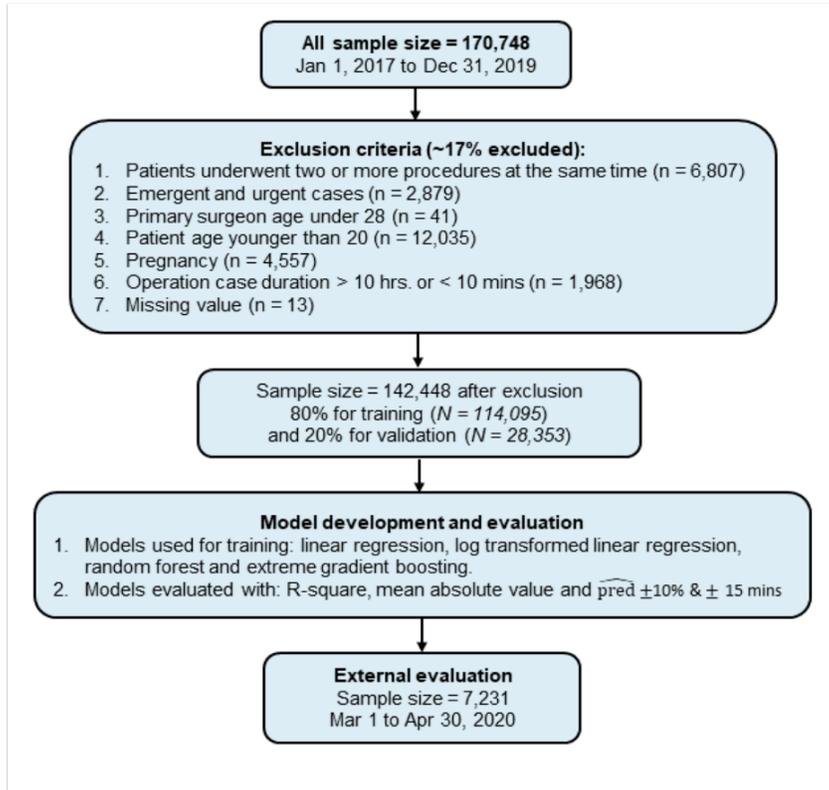


Figure S2: The workflow of model training for this study. The data used for model training fall within the time range of Jan 1, 2017, to Dec 31, 2019. From this data set, approximately 17 % of the cases were excluded based on the following criteria: patients with two or more surgical procedures performed at the same time, emergent and urgent cases, surgeons with age under 28, patients with age younger than 20, pregnant patients, procedure duration longer than 10 hours or less than 10 minutes and cases with missing values. The total number of cases included in the data set for model building was 142,448. This data set was then split into training (80 %) and validation (20 %) subsets for model development. Machine learning and linear regression models were developed on the training data set and validated on the validation data set using R-square and mean absolute error. The percentage of cases with actual duration differences falling within 10 % and 15 minutes of the predicted procedure duration was also computed. Eventually, the models were further evaluated on the most recent surgical cases (from Mar 1 to Apr 30, 2020), which were not included in the original data set for model training.