Artificial Intelligence Generated Analytics from Laparoscopic Video Estimates Operative Experience

Thomas M. Ward, MD [1,2] Daniel A. Hashimoto, MD, MS [1,2] Yutong Ban, PhD [1,3] Guy Rosman, PhD [1,3] Ozanan R. Meireles, MD [1,2]

[1]Surgical AI and Innovation Laboratory, Massachusetts General Hospital [2]Department of Surgery, Massachusetts General Hospital [3]Computer Science and Artificial Intelligence Laboratory, Massachusetts Institute of Technology

Assessing operative experience with advanced automated analytics

Assessing laparoscopic video is a time-intensive process. We have previously trained Artificial Intelligence (AI) models to accurately identify surgical phases in laparoscopic video. [5, 2] These models output, for each second of video, the likelihood and certainty of each operative phase (e.g., clip cystic artery, cut cystic artery). With laparoscopic cholecystectomy as our test procedure, we investigated if the output of these models could be used to automatically assess the video and predict the magnitude of the surgeon's years in practice (MYIP).

We trained an AI model to identify 12 operative phases in 80 laparoscopic cholecystectomy videos. We then tested it on 120 separate videos, recording the likelihood and certainty it predicted for each phase in each second of video. From this model output, we calculated forty-five metrics to quantify performance, such as the duration of phases, the model's certainty, and the number of phase transitions.

We investigated the association of these metrics with MYIP. First, we used a Principle Components Regression (PCR) approach. PCR is a regression technique that uses Principle Components Analysis (PCA) to identify groups of metrics that explain variability in the dataset then estimated these groups' ability to predict MYIP using standard linear regression. This model performed well, with a root mean square error (RMSE) of just 2 years on cross-validation. Second, to obtain estimates for each individual metric, we performed a multi-variable analysis to estimate MYIP from all the calculated metrics. Due to the large number of predictors, we used a lasso-penalized regression to select the inputs that explained the greatest variability in training year, using the lasso penalty that minimized mean cross-validated error.[1] Post-selection inference to derive p-values on the lasso-selected variables was performed using the *selectiveInference* package.[3] All calculations were performed using R version 4.0.5.[4]

Majority of surgeon variability is explained by the number of times they switch between phases and their operative duration

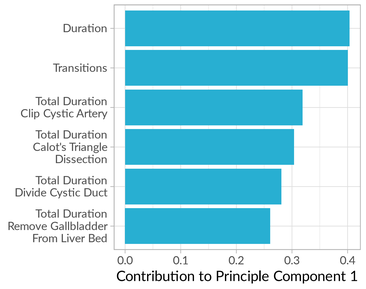
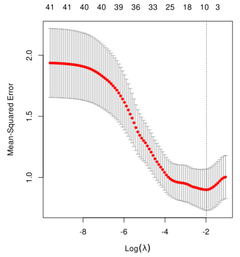


Figure 1. The graph shows the contribution (weight that each value was multiplied by) for the six most important metrics that composed Principle Component 1 (PC1) in the PCR. As in all PCR, PC1 provides the line that most closely fits the inputs. Therefore, knowing the duration of the major operative phases (Calot's Triangle Dissection and Removal of the Gallbladder from the Liver Bed) and the case overall, combined with the number of phase transitions (the number of times the surgeon flipped between performing one phase to another) would allow you to best discriminate between the videos. In the PCR, PC1 was not substantially associated with MYIP (p = 0.99), which means that knowing how long the case took did not help with predicting MYIP.

Six advanced operative metrics predict magnitude of years in practice

Cholecystectomy videos (N = 120) from 19 surgeons, each second labeled with an operative phase by an AI model.

AI-Generated Metrics (N = 45):
1. Average phase duration (N = 12)
2. Total phase duration (N = 13)
3. Number phase transitions (N = 2)
4. Phase Certainty (N = 13)
5. Phase ordering around clipping/cutting (N = 5)



Lasso regression:
log(Years in Practice) ~ metrics

Metric	Coefficient	p-value
Certainty Clip Cystic Duct	0.25	0.003
Mean Duration Calot's Dissection	-0.26	0.004
Clip Duct First	0.16	0.02
Mean Duration Release GB Peritoneum	-0.23	0.9
Total Duration Clip Cystic Duct	0.08	0.9
Total Duration Calot's Dissection	-0.24	0.9

Adjusting for the level of gallbladder (GB) inflammation (Parkland Grading Scale (PGS)), lasso-penalized regression of MYIP selected 6 of the original 45 metrics for the highest-performing model. Post-selection inference showed, after adjustment for other selected covariates, that more experienced surgeons tended to clip the cystic duct first, dissect the triangle of Calot more quickly, and dissect out the cystic duct in a manner such that the AI model could recognize the action of clipping with higher certainty. This suggests that more experienced surgeons tend to do a better job at isolating the cystic duct and clip it first, without needing to clip and divide the cystic artery to facilitate their dissection.

Younger attendings perform longer dissections and dissect out all cystic structures prior to clipping/cutting

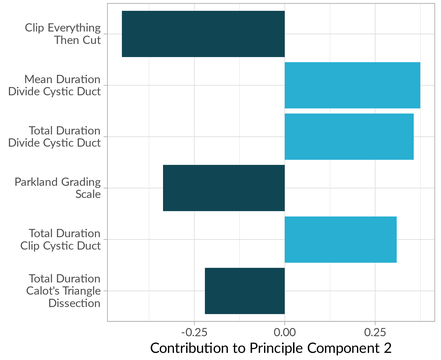


Figure 2. Principle Component 2 (PC2) in the PCR had a significant positive correlation with MYIP (p = 0.02). More experienced surgeons were faster at Calot's triangle dissection and less likely to clip all structures before dividing them, suggesting that they relied more on pattern recognition to, earlier in the case, clip and divide one cystic structure before moving towards dissection and division of other cystic structures. Less experienced surgeons, meanwhile, took longer to dissect, then clipped and divided all structures at the same time. Some of this variation may be due to younger surgeons having more inflamed gallbladders, as seen by the negative contribution of PGS to PC2. Interestingly, older surgeons took longer during the actually clipping process, which may reflect their comfort with allowing younger trainees to perform these potentially more dangerous phases.

Implications for automated operative assessment

- Automated intra-operative metrics correlate with surgeons' clinical experience.
- Personalized metrics could help target areas of weakness to accelerate development of operative skills.

Future directions

- Correlate metrics to visual assessments of operative performance.
- Apply automated metrics to other laparoscopic cases

References

[1] Jerome Friedman et al. *Journal of statistical software*. 2010.
[2] Daniel A. Hashimoto et al. *Annals of Surgery*. Sept. 2019.
[3] Jason D. Lee et al. *Annals of Statistics*. June 2016.
[4] R Core Team. 2020.
[5] Thomas M. Ward et al. *Surgical Endoscopy*. July 27, 2020.