Assessment of Surgical Learning Curves in Transoral Robotic Surgery for Squamous Cell Carcinoma of the Oropharynx

William G. Albergotti, MD,

Department of Otolaryngology–Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

William E. Gooding, MS,

Biostatistics Facility, University of Pittsburgh Cancer Institute, Pittsburgh, Pennsylvania

Mark W. Kubik, MD,

Department of Otolaryngology–Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Mathew Geltzeiler, MD,

Department of Otolaryngology–Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Seungwon Kim, MD,

Department of Otolaryngology–Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Umamaheswar Duvvuri, MD, PhD, and

Department of Otolaryngology–Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Veterans Affairs Pittsburgh Health System, Pittsburgh, Pennsylvania

Robert L. Ferris, MD, PhD

Department of Otolaryngology–Head and Neck Surgery, University of Pittsburgh Medical Center, Pittsburgh, Pennsylvania

Abstract

IMPORTANCE—Transoral robotic surgery (TORS) is increasingly employed as a treatment option for squamous cell carcinoma of the oropharynx (OPSCC). Measures of surgical learning curves are needed particularly as clinical trials using this technology continue to evolve.

OBJECTIVE—To assess learning curves for the oncologic TORS surgeon and to identify the number of cases needed to identify the learning phase.

DESIGN, SETTING, AND PARTICIPANTS—A retrospective review of all patients who underwent TORS for OPSCC at the University of Pittsburgh Medical Center between March 2010 and March 2016. Cases were excluded for involvement of a subsite outside of the oropharynx, for nonmalignant abnormality or nonsquamous histology, unknown primary, no tumor in the main specimen, free flap reconstruction, and for an inability to define margin status.

EXPOSURES—Transoral robotic surgery for OPSCC.

MAIN OUTCOMES AND MEASURES—Primary learning measures defined by the authors include the initial and final margin status and time to resection of main surgical specimen. A cumulative sum learning curve was developed for each surgeon for each of the study variables. The inflection point of each surgeon's curve was considered to be the point signaling the completion of the learning phase.

RESULTS—There were 382 transoral robotic procedures identified. Of 382 cases, 160 met our inclusion criteria: 68 for surgeon A, 37 for surgeon B, and 55 for surgeon C. Of the 160 included patients, 125 were men and 35 were women. The mean (SD) age of participants was 59.4 (9.5) years. Mean (SD) time to resection including robot set-up was 79 (36) minutes. The inflection points for the final margin status learning curves were 27 cases (surgeon A) and 25 cases (surgeon C). There was no inflection point for surgeon B for final margin status. Inflection points for mean time to resection were: 39 cases (surgeon A), 30 cases (surgeon B), and 27 cases (surgeon C).

CONCLUSIONS AND RELEVANCE—Using metrics of positive margin rate and time to resection of the main surgical specimen, the learning curve for TORS for OPSCC is surgeon-specific. Inflection points for most learning curves peak between 20 and 30 cases.

Since its first described surgery in patients in 2008, transoral robotic surgery (TORS) for the resection of oropharyngeal squamous cell carcinoma (OPSCC) has increasingly gained use.[1,2] It has been rapidly adopted owing to its reported excellent oncologic outcomes, hope for improved long-term functional outcomes compared with chemoradiation and the possibility for de-escalated adjuvant radiotherapy.[2-4] With this rapid adoption of a new surgical technology, there has been increasing need on the part of hospital and clinical trial credentialing committees to determine adequacy of training and experience of TORS surgeons.[5] A recent guideline from the 2015 American Head and Neck Society (AHNS) Education Committee, American Academy of Otolaryngology–Head and Neck (AAO-HNS) Surgery Robotic Task Force, and AAO-HNS Sleep Disorders Committee provided a reference for institutions charged with training and credentialing robotic head and neck surgeons but still noted the wide variety in credentialing practices across the country.[5] It suggested a minimum of 20 cases (10 bedside/10 console) for graduates of a residency or fellowship program with TORS training or 2 proctored cases for those surgeons completing

approved training courses. The NCI-funded ECOG 3311 trial, which is the largest of several prospective trials evaluating the oncologic efficacy of TORS, requires surgeons to have completed 20 oncologic TORS cases prior to credentialing, among other requirements.[4] Recently, the ECOG 3311 study team reported a positive margin rate of approximately 3% suggesting that credentialing using a minimum experience of 20 cases was reasonable; however, there is little objective evidence supporting this number.[6] Unfortunately, data on learning curves in TORS for OPSCC is lacking, leaving decisions to be made primarily based on expert opinion.

The theory of the learning curve is predicated on worse performance at the start of training, followed by better performance with experience. A learning curve that reflects this theoretical learning process would rise initially, and then fall as the trainee acquires skill and performance improves. Published reports of learning curves for TORS have focused on robotic set-up time or time to resection and range from 20 to 42 cases, with consensus suggesting that 20 cases tends to herald the end of a learning period.[7-9] These studies included both benign and malignant tumors as well as nonoropharyngeal subsites, limiting their interpretability. In addition, no reports could be identified that evaluated margin status, which would be of great interest to an oncologic TORS surgeon, particularly in the effort to avoid "3 modality" therapy because a positive margin would generally mandate high-dose chemoradiotherapy (CRT), obviating upfront surgery in good prognosis cases. Cumulative sum (CUSUM) learning curves have been developed to evaluate procedural failures over time, such as a positive margin, and have been employed previously for the study of surgical learning and quality monitoring in general and in robotic surgery in several surgical fields.[10-12] An individual surgeon's performance over time can therefore be measured by comparing individual outcomes against an internal standard (his own overall average performance) or against an external standard supported by literature, such as an acceptable positive margin or complication rate. The CUSUM is an accepted metric for evaluating learning effects.[13] We therefore designed this retrospective study to investigate the learning curves regarding tumor margin status of 3 experienced oncologic TORS surgeons. We hypothesized that evidence of learning would be seen at 20 cases, based on the existing literature.

Methods

This is a retrospective cohort study of all TORS cases performed for oropharyngeal squamous cell carcinoma at the University of Pittsburgh Medical Center between March 2010 and March 2016. Surgeons A and C were active for the entirety of the study period, whereas surgeon B was credentialed from March 2011 to March 2016. All procedures were performed with either the DaVinci Si or Xi system (Intuitive Surgical, Sunnyvale, CA). Approval for this study was obtained from the University of Pittsburgh institutional review board. Informed consent was waived owing to the retrospective nature of this study. The surgery schedule was searched for all cases of TORS of the oropharynx. Cases were excluded if part of the surgery involved a subsite of the upper aerodigestive tract outside of the oropharynx, for nonmalignant abnormality, for nonsquamous histology, unknown primary at the time of surgery, no tumor in the main oropharyngeal specimen, surgery requiring free flap reconstruction, and for an inability to define margin status (ie, only

diagnostic procedure). Three surgeons performed all procedures as received through referral patterns. Baseline patient characteristics and postoperative course data was obtained through medical record review. Cases were stratified by surgeon and ordered chronologically.

Analyses of learning curves were conducted using the CUSUM technique. In this method a constant failure rate is assumed over the observation period such that each successful case causes the learning curve to be adjusted downward by an amount representing the expected failure rate. A failure is denoted by a rise in the CUSUM curve. If on average the number of failures agrees with the expected rate the curve will be flat. An increase denotes learning is occurring until the curve either flattens or turns downward.

The primary metric selected for evaluating competency in robotic TORS oropharyngeal surgery was final margin status. Secondary competency measures included initial margin status (ie, initial main specimen or surgical field margins), time to resection, duotube placement rate, hospital length of stay, and 30-day readmissions. Bleeding rates were not included as a measure of learning because halfway through this series all 3 surgeons began ligating branches of the external carotid to attempt to reduce severe bleeding complications of TORS.[14] Thus it was not felt that it would be an appropriate measure of surgical learning curve because it was a new surgical procedure. Margin status was determined after review of the pathology reports by the primary author and the operating surgeon with agreement on final margin status by both parties. Time to resection was defined as time from initiation of robot docking to resection of the main specimen. This included time for marking of the specimen.

Owing to similarity among risk profiles for each surgeon, the individual surgeon average was chosen as the reference standard. Categorical metrics margin and 30-day readmission used the surgeon average over the series. The continuous metrics, time to resection and hospital length of stay, used the surgeon individual mean and standard deviation. The computation methods used for the analysis can be found in Grunkemeier et al.[15]

Additional statistical methods included the comparison of baseline patient characteristics by surgeon using a χ^2 or Cochran-Armitage trend test for categorical variables and a Kruskal-Wallis test for continuous variables.

Results

During the study period, there were 382 transoral robotic procedures identified. Cases were excluded for involvement of the nonoropharyngeal subsites of the upper aerodigestive tract (39 cases), nonmalignant abnormality (71 cases), nonsquamous histology (13 cases), unknown primary (68 cases), no tumor in the main oropharyngeal specimen (12 cases), inability to define margin status (13 cases), and free flap reconstruction (6 cases). This left 160 cases (125 men and 35 women with a mean [SD] age of 59.4 [9.5] years) that met our inclusion criteria: 68 for surgeon A, 37 for surgeon B, and 55 for surgeon C. Of 160 tumors, 140 (87.5%) were T1 or T2 and 126 (78.8%) were human papillomavirus positive. The most common subsite was the tonsil (57.5%), followed by the base of tongue (36.3%). There were no revision TORS cases, although 13 (8.1%) of the 160 had been previously treated with

chemoradiotherapy. There were no differences in the baseline patient characteristics between the 3 surgeons except the median patient age (55 years [interquartile range (IQR) 49-61] for surgeon B vs 61 [IQR 54-66] and 60 years [IQR 55-67] for the other 2 surgeons, P = .02 (Table). Mean (SD) time to resection including robot set-up was 79 (36.9) minutes. There was no statistically significant difference in time to resection between the 3 surgeons. Of 159 patients, 29 (18.2%) required a temporary feeding tube in the immediate postoperative period. Twenty-four (15.1%) patients were readmitted within 30 days of surgery. The median length of stay for the initial hospitalization was 2 days (range, 1-24 days).

The overall final positive margin rate was 22 (13.7%) of 160 and was similar between the 3 surgeons (surgeon A, 14.7%; surgeon B, 13.5%;surgeon C, 12.7%). Inflection points for final margin status were 27 cases (surgeon A) and 25 cases (surgeon C) (Figure 1). There was no inflection point identified for final margin status for surgeon B. Inflection points for initial margin status were 15 cases (surgeon B) and 22 cases (surgeon C) (Figure 2). There was no inflection point identified for initial margin status for Surgeon A. Inflection points for mean time to resection were 40 cases (surgeon A), 30 cases (surgeon B), and 27 cases (surgeon C) (Figure 3). There was no consistent evidence of a learning curve for duotube placement rate, hospital length of stay, or 30-day readmissions. To investigate whether experience and skill acquisition was commensurate with increased case difficulty, we examined temporal change in the proportion of base of tongue vs tonsil or distribution of T stage. There was no significant change in either measurable variable over time suggesting no evidence of risk shift or more difficult cases as the surgeon progresses in skill and comfort with the instrumentation.

Discussion

Learning curves have been used in the medical literature both to demonstrate evidence of skill acquisition and competence, which would be anticipated as an operator gains experience, and also to show changes in performance over time. This assumes worse performance at the start of a new technique (soon after training), followed by better performance with experience, higher case numbers, and greater complexity. This is the first study to demonstrate evidence for a learning curve for margin status after TORS. In this study of learning curve for TORS for OPSCC we found that the CUSUM learning curve is variable by surgeon but ranges between 15 to 40 cases depending on the metric chosen, with most learning curves showing inflection points by 30 cases. This is longer than our hypothesized inflection point of 20 cases. A particularly good example of a learning curve in our series is final margin status for surgeon C, which shows a cluster of positive margins early in his TORS experience followed by fewer positive margins later on (and none after case 40, which may be owing to acquisition of surgical skill or a lucky run of cases). However, the inflection point is not a static number and could change as surgeon experience grows.

Previous studies have suggested that learning curves for TORS typically peak between 20 to 42 cases. Genden et al[8] published his first 20 cases, and of the 18 cases which were taken to completion, he showed improvement in robot setup time by the end of his series. Resection times were not reported but both benign and malignant tumors were included, which would

make resection times more difficult to interpret because cases of benign tumors have been reported by others to have significantly shorter resection times.[3]

White and Magnuson[7] reported their first 168 TORS cases, divided them into chronological quartiles (42 patients each), and reported significant improvement in operative time, duration of intubation, and hospital stay after the first 42 cases. While cases of both benign and malignant tumors were included, there was an equal distribution of these cases in each quartile. Margin status was not reported in this series.

O'Malley and Weinstein[9] suggest that 20 is the number of cases needed to reach the peak of a learning curve. Using themselves as a reference standard, they have found that neophyte TORS surgeons tend to reach the operative times of an expert after about 20 cases. They also note that margin status does not tend to improve over time, although these data are not shown. Given the limited evidence in the literature, a recent set of guidelines by the 2015 AHNS Education Committee, AAO-HNS Surgery Robotic Task Force, and AAO-HNS Sleep Disorders Committee recommended at least 20 cases (divided between bedside and console) for surgeon training in residency fellowship and a full training course, observation of an experienced TORS surgeon, 2 cadaver dissections, and 2 proctored TORS cases for those surgeons already in practice; however, they noted that their recommendations were based on limited evidence in the literature.[5]

While time to resection is an important measure of surgeon competence we feel that margin status is of greater importance to an oncologic TORS surgeon. Clinical trials, in particular, need to be able to credential surgeons who have reached the endpoint of their learning curve. For instance in ECOG 3311, the goal of the trial is to ask whether postoperative radiotherapy could be reduced (deintensified) in good to intermediate prognosis cases, where R0 surgical resection had been rendered upfront. Thus, "learning" on the trial by each surgeon was deemed unacceptable, and a novel credentialing process was designed and implemented. Of more than 100 surgeons who applied, approximately 75 active surgeons have been credentialed. This has led to accrual of over 325 case with an aggregate positive margin rate of approximately 3%.[6]

In this study we found final margin status to be a valid measure of surgeon competence with a peak for 2 of the 3 surgeons between 25 to 27 cases. The third surgeon (surgeon B) has a relatively flat learning curve which may be owing to that surgeon's lower number of cases, and with more cases there may be a more clear inflection point to his learning curve. His learning curve for initial margin status has an inflection point at 15 cases. Surgeon A had a flat learning curve for initial margin status while surgeon C's learning curve had an inflection point at 22 cases, similar to his inflection point for final margin status. Each of the 3 surgeons had a clear inflection point for time to resection with all but 1 (surgeon A) reaching that point by 30 cases. Therefore, we conclude that evidence of learning using these 3 metrics can be seen for at least 2 of 3 surgeons and that most of these inflection points occur between 20 and 30 cases. Of note, our positive margin rate (13.7%) is higher than others have reported in the literature, which generally ranges from 5% to 10%.[1,3,16] We feel this is likely owing to the inclusion of an unselected group of patients and our reporting as an intention to treat (ie, all patients with an attempt at resection are included).

In addition to final margins and time to resection we investigated the utility of surgical field margins, length of hospital stay, and 30-day readmission rate as potential learning metrics. None of these additional metrics were useful and may be insensitive to or overly crude indicators of skill acquisition. Longer surgical series may be needed to demonstrate whether they have learning curve utility.

A posteriori, each of the individual surgeons was asked when subjectively they felt that they had reached the peak of their surgical learning curve. The times were variable: surgeon A answered 10 cases, surgeon B was unable to identify a specific time point, and surgeon C answered 30 cases. It would be valuable for future studies to ask this question prior to data analysis to gauge accuracy of surgeon intuition.

Limitations

There are several limitations of this study. Training or proctored cases were excluded. Each of these surgeons were trained in TORS after their residency/fellowship training and these results may not be applicable to those with experience during their primary training. In addition, each of these surgeons is an experienced head and neck surgeon with several years of open resection experience prior to attempting TORS. Furthermore our institution is a tertiary academic referral center and cases may be atypical compared with institutions offering TORS. We used each surgeon's overall positive margin rate as an internal standard, therefore assuming that these surgeons have reached competency by the end of their learning curves. It may be that if applied to an external standard (accepted positive margin rate) the results would be different. However, we do not feel that such a standard exists currently in the TORS literature. The primary endpoint, final margin status, was defined with input from the individual surgeon and thus is not an independent metric.

A strength of this study is that cases of benign tumors were not included, allowing us to focus exclusively on oncologic cases. Cases of benign tumors were excluded because margin status is not relevant and because time of resection has been shown to be considerably lower in these cases. However, it may be that surgeons master benign TORS before moving on to oncologic TORS. Another feature of this study is that we investigated whether, as learning progresses, the surgeon may be more willing to take on riskier procedures thereby dynamically abrogating quality improvement metrics. Although we did not see evidence for a shift to higher T classification tumors or more BOT tumors over time, it is possible that there is a shift to riskier procedures, therefore shifting the learning curve to the right.

We would advocate for publication of other surgeons' learning curves to better determine the range of normal for TORS for OPSCC and to develop standardized means. In addition, learning curves such as these reported should be maintained going forward as quality control mechanisms to evaluate surgeons over time for better or worse performance.

Conclusions

Using metrics of positive margin rate and time to resection of the main surgical specimen, the learning curve for TORS for squamous cell carcinoma of the oropharynx is

surgeonspecific and ranges between 15 and 40 cases. Inflection points for most learning curves peak between 20 and 30 cases.

Acknowledgments

References

1. Weinstein GS, O'Malley BW Jr, Snyder W, Sherman E, Quon H. Transoral robotic surgery: radical tonsillectomy. Arch Otolaryngol Head Neck Surg. 2007; 133(12):1220–1226. [PubMed: 18086963]

2. Chen MM, Roman SA, Kraus DH, Sosa JA, Judson BL. Transoral Robotic Surgery: A Population-Level Analysis. Otolaryngol Head Neck Surg. 2014; 150(6):968–975. [PubMed: 24618503]

3. Weinstein GS, O'Malley BW Jr, Magnuson JS, et al. Transoral robotic surgery: a multicenter study to assess feasibility, safety, and surgical margins. Laryngoscope. 2012; 122(8):1701–1707. [PubMed: 22752997]

4. Holsinger FC, Ferris RL. Transoral endoscopic head and neck surgery and its role within the multidisciplinary treatment paradigm of oropharynx cancer: robotics, lasers, and clinical trials. J Clin Oncol. 2015; 33(29):3285–3292. [PubMed: 26351337]

5. Gross ND, Holsinger FC, Magnuson JS, et al. Robotics in otolaryngology and head and neck surgery: recommendations for training and credentialing: A report of the 2015 AHNS education committee, AAO-HNS robotic task force and AAO-HNS sleep disorders committee. Head Neck. 2016; 38(suppl 1):E151–E158. [PubMed: 26950771]

6. Ferris RL, Holsinger S, Weinstein FC, et al. E3311 trial of transoral surgery for oropharynx cancer: Implementation of a novel surgeon credentialing and quality assurance process. J Clin Oncol. 2016; 34(suppl:abstr 6054)

7. White HN, Frederick J, Zimmerman T, Carroll WR, Magnuson JS. Learning curve for transoral robotic surgery: a 4-year analysis. JAMA Otolaryngol Head Neck Surg. 2013; 139(6):564–567. [PubMed: 23680949]

8. Genden EM, Desai S, Sung CK. Transoral robotic surgery for the management of head and neck cancer: a preliminary experience. Head Neck. 2009; 31(3):283–289. [PubMed: 18972413]

9. Weinstein, GS., O'Malley, BW. TransOral robotic surgery (TORS). San Diego: Plural Pub; 2012.

10. Barrie J, Jayne DG, Wright J, Murray CJ, Collinson FJ, Pavitt SH. Attaining surgical competency and its implications in surgical clinical trial design: a systematic review of the learning curve in laparoscopic and robot-assisted laparoscopic colorectal cancer surgery. Ann Surg Oncol. 2014; 21(3):829–840. [PubMed: 24217787]

11. Cundy TP, Gattas NE, White AD, Najmaldin AS. Learning curve evaluation using cumulative summation analysis-a clinical example of pediatric robot-assisted laparoscopic pyeloplasty. J Pediatr Surg. 2015; 50(8):1368–1373. [PubMed: 25783405]

12. Novick RJ, Stitt LW. The learning curve of an academic cardiac surgeon: use of the CUSUM method. J Card Surg. 1999; 14(5):312–320. [PubMed: 10875583]

13. Williams SM, Parry BR, Schlup MM. Quality control: an application of the cusum. BMJ. 1992; 304(6838):1359–1361. [PubMed: 1611337]

14. Mandal R, Duvvuri U, Ferris RL, Kaffenberger TM, Choby GW, Kim S. Analysis of post-transoral robotic-assisted surgery hemorrhage: Frequency, outcomes, and prevention. Head Neck. 2016; 38(suppl 1):E776–E782. [PubMed: 25916790]

15. Grunkemeier GL, Wu YX, Furnary AP. Cumulative sum techniques for assessing surgical results. Ann Thorac Surg. 2003; 76(3):663–667. [PubMed: 12963172]

16. Blanco RG, Fakhry C, Ha PK, et al. Transoral robotic surgery experience in 44 cases. J Laparoendosc Adv Surg Tech A. 2013; 23(11):900–907. [PubMed: 24083851]

Key Points

Question

This study assesses learning curves for the oncologic transoral robotic surgery (TORS) surgeon and elucidates the number of cases needed to identify the learning phase.

Results

In this analysis learning curves for TORS for oropharyngeal cancer were surgeon-specific but tended to show inflection points between 20 and 30 cases.

Meaning

Evidence for learning should be tracked for surgeons using TORS as an oncologic treatment modality, which may be used as a guide to assist with surgeon credentialing.
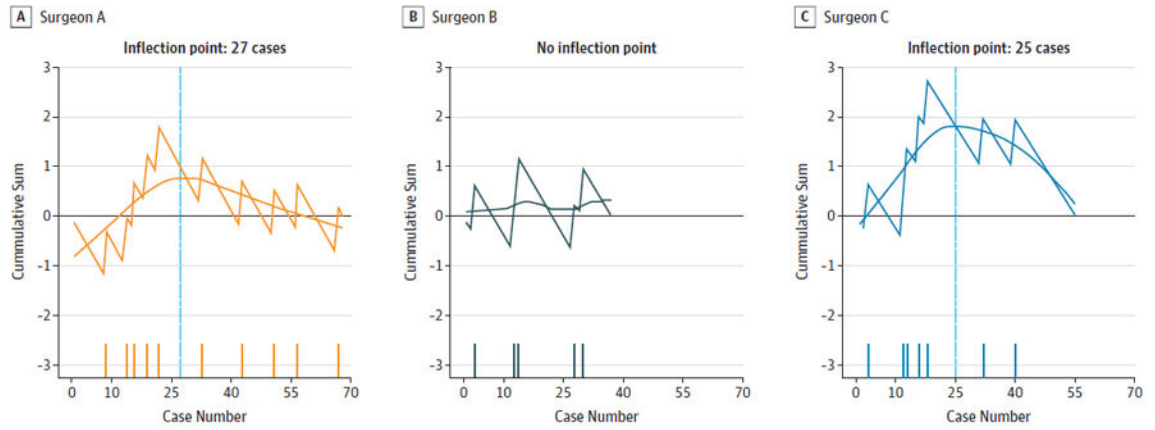
Figure 1. Positive Final Margin Cumulative Sum Learning Curve, Stratified by Surgeon
A-C, Surgeons A, B, and C. The horizontal reference at 0 denotes the individual final average. The *y*-axis denotes cumulative sum of the expected failure proportion from the series mean at each time point. The *x*-axis denotes consecutive cases. The blue vertical line indicates the inflection point. The dashes along the *x*-axis indicate positive margin cases.
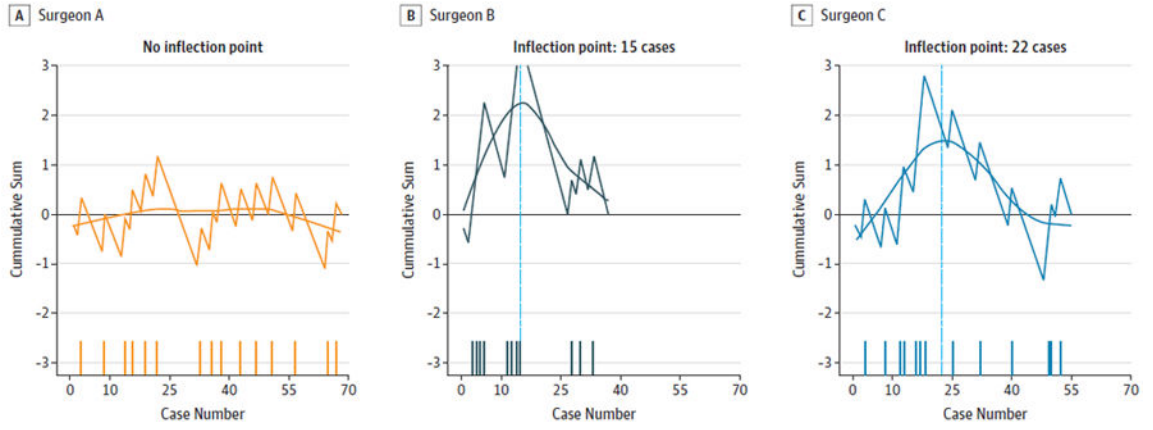
Figure 2. Positive Initial Margin Cumulative Sum Learning Curve, Stratified by Surgeon
A-C, Surgeons A, B, and C. The horizontal reference at 0 denotes the individual average. The y-axis denotes cumulative sum of the expected failure proportion from the series mean at each time point. The x-axis denotes consecutive cases. The dashes along the x-axis indicate positive margin cases.
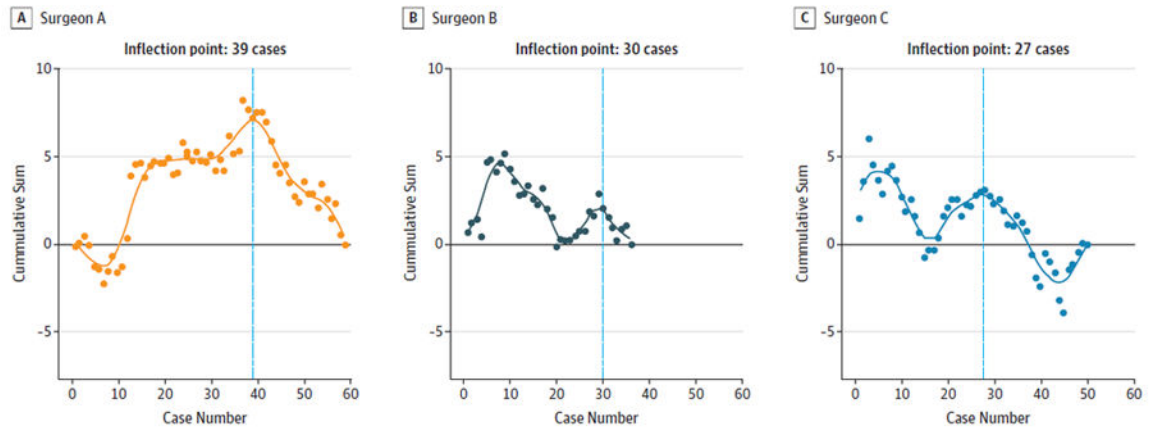
Figure 3. Time to Resection of Main Tumor Specimen (Including Robot Docking and Set-up Time), Stratified by Surgeon

A-C, Surgeons A, B, and C. The *y*-axis denotes cumulative sum of the standard deviation from the series mean at each time point. The *x*-axis denotes consecutive cases.

Table

Baseline Characteristics of Included Patients, Stratified by Surgeon

Characteristic	Surgeon A	Surgeon B	Surgeon C	P Value
Patients, No.	68	37	55	
Age, mean (IQR), y	61 (54-66)	55 (49-61)	60 (55-67)	.02
Sex				.76
Male	54	28	43	
Female	14	9	12	
Site				.22
Base of tongue	22	10	26	
Tonsil	4	25	25	
Other	42	2	4	
pT stage				.10
1	23	16	24	
2	31	21	25	
3	12	0	5	
4	0	0	1	
T×	2	0	0	
pN Stage				.31
0	16	9	10	
1	9	12	13	
2	38	15	26	
3	1	0	2	
T×	4	1	4	
Indication				.95
Primary	62	34	51	
Recurrence	6	3	4	
P16				.21
Positive	51	33	42	
Negative	17	4	13	

Abbreviations: IQR, interquartile range; pT, primary tumor; pN, lymph node metastases; T×, unknown primary.