Leader Brittany (Orcid ID: 0000-0001-5555-2046)
Ishman Stacey (Orcid ID: 0000-0003-0997-9692)
Skinner Margaret (Orcid ID: 0000-0002-6305-3981)
Wiet Gregory (Orcid ID: 0000-0001-6990-3206)

Validation of an Objective Assessment Tool for Tonsillectomy in Training

Short Title: Validation of a Tonsillectomy OS

*Luke Jakubowski, MD[1], *Brittany A. Leader, MD[2], Stacey L. Ishman Chun, MD[5], Sergey S. Tarima, PhD[6], Sanjay Parikh, MD[7], Margaret S Thorne, MD[9], Robert Weatherly, MD[10], Gregory Wiet, MD[11], David J
*Co-first authors

[1]Department of Otolaryngology-Head & Neck Surgery, University of Minne USA.
[2]Department of Otolaryngology-Head & Neck Surgery, University of Cincin Cincinnati, OH, USA.
[3]Division of Pediatric Otolaryngology, Cincinnati Children's Hospital Medic USA.
[4]Division of Pulmonary Medicine, Cincinnati Children's Hospital Medical C
[5]Department of Otolaryngology & Communication Sciences, Medical Colleg WI, USA.
[6]Division of Biostatistics, Institute for Health and Equity, Medical College WI, USA.
[7]Department of Otolaryngology -Head and Neck Surgery, University of Was
[8]Department of Otolaryngology -Head and Neck Surgery, Johns Hopkins Sc MD, USA.
[9]Department of Otolaryngology -Head and Neck Surgery, University of Mic
[10]Department of Otolaryngology -Head and Neck Surgery, University of Kar
[11]Department of Otolaryngology -Head and Neck Surgery, Nationwide Child State University, Columbus, OH, USA.

Corresponding Author:

Email: Stacey.Ishman@cchmc.org

Financial Disclosure: The authors have nothing to disclose.
Conflict of Interest: The authors have no conflicts of interest.
Key Words: objective structured assessment of technical skills, OSAT
resident education, tonsillectomy

Abstract

Objective: Create and validate an objective structured assessment of t
for otolaryngology residents learning how to perform a tonsillectomy.

Study Design: Multicenter prospective longitudinal validation study

Methods: A multi-institutional study at six tertiary academic otolaryn
from 7/09-5/12. Using the modified Delphi technique, a panel of pedi
created a tonsillectomy task-based checklist (TBC) for a tonsil OSATS
type scale. Residents were assessed by pediatric otolaryngology staff
the TBC and a global rating scale (GRS). Procedure time, patient age
performed tonsillectomies, and surgical technique were also collected.

Results: One-hundred-sixty-seven tonsil OSATS were completed for
competency was recorded for 99 (59.2%). Residents scored as compe
significantly more previous tonsillectomies than those deemed non-co
13.5±11.6, respectively (*P<0.001)*. The mean overall score on the ton
2.6±1.0 for competent and non-competent, respectively (*P<0.001)*. H
tonsillectomies performed and mean tonsil TBC score significantly inc
competency (*P*<0.001). Each additional tonsillectomy performed incr
achieving competency by 6.3% (*P*=0.006; 95%CI:1.330-1.110), and e
mean tonsil TBC score increased the likelihood of competency by a fa

Level of Evidence: 4

Introduction

The training and surgical skills assessment models introduced by Hals
were based upon subjective evaluation. In the 2001 Outcomes Project,
(Accreditation Council for Graduate Medical Education) mandated tha
objectively assess resident surgical skills[1]. However, surgical residenc
of achieving this mandate[2].

The ACGME developed and implemented the Outcomes Project with
improve resident education through integration of 6 core competencie
assessment.[1] Since the ACGME mandate, a number of surgical specia
implemented objective assessment tools to measure competency such
Assessment of Technical Skills (OSATS). OSATS is a framework dev
which consists of both a task-based checklist (TBC) and global rating
must be developed for each procedure needing assessment and althoug
use of this approach does not in and of itself imply validity or reliabili
however, OSATS developed for procedures in other surgical fields ha
validity, reliability, and feasibility[4]. Unfortunately, Otolaryngology sp
used among 15.3 % of otolaryngology residency programs.[2]

OSATS assessments have been praised for their proven track record, h
studies were performed in a laboratory setting.[5-9] In 2010, van Hove *e*
objective assessment of surgical skills and only 28 (26.9%) were perfo
room.[5] Similarly, while there has been increasing interest in OSATS i
been more frequently used in the laboratory setting. The use of simula
are especially beneficial for learning emergency or more complex proc
airway foreign body removal[10], transoral robotic surgery[11], endoscopic
mastoidectomy.[14] However, it is important to develop and validate ass
the intraoperative setting as well. Therefore, single centers have begu
assessment tools.[15,16]

Currently, most otolaryngology training programs evaluate surgical pe
feedback provided at the end of the month or rotation, feedback which

Tonsillectomy is one of the most common surgical procedures perform
residency training. US otolaryngology residents perform an average o
during their training[17]. However, the number of tonsillectomies perfo
competency has not been studied. Understanding when competency is
allow the resident to redirect their time and efforts towards mastery of
or those where they have not yet achieved competency. This is particu
onset of residency work hour restrictions which limit clinical exposure
efficiency in graduate medical education.

With the ACGME outcomes mandate in mind, the primary aim of this
was to develop and validate an objective assessment tool for tonsillect
formative feedback, prediction of competency, with the ultimate goal
resident time when learning surgical procedures. We hypothesized tha
performed more tonsillectomies were more likely to be competent, an
objective assessment of their skills. To our knowledge this represents
tonsillectomy OSATS study performed in the operating room.

Methods

Using the modified Delphi technique, an OSATS evaluation tool for to
OSATS) was developed in 2009 by a panel of pediatric otolaryngolog
institutions: Medical College of Wisconsin, University of Michigan S
Hopkins School of Medicine, The Ohio State University College of M
Kansas Medical School, and University of Washington School of Med
consists of a TBC along with a subjective rating of overall tonsillector
[**Figure 1**].

The panel identified 10 pertinent steps in performing tonsillectomy an
the Tonsillectomy TBC. Each step was evaluated on a five-point Likert-type
'Not Performed' rating. 'Not Performed' identified aspects of the proc

Many studies regarding surgical education use the postgraduate year (advanced a resident should be in their training. Because the period wh primarily taught can vary by program, we used the absolute number of predictor of competence in this study.

In addition, patient age (in years), and the tonsillectomy method were electrocautery, or other). Immediately after the procedure, the staff pe completed the Tonsil TBC. Scores were based on the case just comple feedback on concomitant procedures such as adenoidectomy or ear tub to scoring the objective OSATS, the staff pediatric otolaryngologist su resident's competency with tonsillectomy GRS, scored as a yes or no, they just completed [**Figure 1**]. As there can be challenging cases for pediatric otolaryngologists were also asked to rate the case complexity could be performed by a general otolaryngologist (hereafter referred t "challenging case for most general otolaryngologists" (hereafter refer final TBC score is reported as a mean of all 10 task scores with a max Tasks rated as 'Not Performed' or missing ratings were not factored in otolaryngologists were given the option of completing the Tonsil OSA password protected web-based version.

All participating institutions obtained Institutional Review Board appr otolaryngology residents on the pediatric otolaryngology service and s otolaryngology faculty at the 6 participating institutions were invited t Cases represent a convenience sample by participating staff pediatric o

Inter-rater reliability, which was defined as a measure of agreement be otolaryngologists evaluating the same procedure, was assessed to furth OSATS. Five cases performed at the Medical College of Wisconsin, a Michigan were selected (as a convenience sample) to evaluate inter-ra faculty were available. For those selected, a tonsillectomy OSATS wa

Statistical Analysis

Analysis of the predictors of tonsillectomy competency, as well as resu
technique, were adjusted for resident effect using Generalized Estimat
the exception of the variables with small cell counts where the Chi-squ
continuity correction was used. The adjusted odds ratios are reported f
ratings of competence for residents evaluated using the OSATS. In ad
intervals were reported from the multiple logistic regression model reg
procedures, tonsillectomy TBC and tonsillectomy technique fitted with
between tonsillectomy procedures performed by inexperienced (<10) a
residents were assessed with Chi-square test with Yates's continuity c
used to compare results with the tonsil TBC. Predictive properties of t
with a Receiver Operating Characteristic curve and area under the cur
assessed on the small group (n=5) of patients who underwent parallel a
0.05 were deemed significant. R version 3.3.2 was used for the analys

Results

A total of 167 Tonsil OSATS were submitted by 14 attendings from 6
assessment of 38 trainees. The number of tonsillectomy procedures, T
tonsillectomy technique were summarized by competency for all 167 s
age was reported for 162 evaluations; the mean was 5.7 ± 3.5 years. D
completed for 163 evaluations; 154 were rated as standard, and 9 were

All had complete data for subjective GRS of competency. Reported su
competency rated 99 (59.2%) of the residents competent to perform a
independently, and 68 (40.7%) were deemed non-competent [**Table 1**
had performed significantly more previous tonsillectomy procedures (
deemed non-competent (13.5 ± 11.6, $P<0.001$). The mean overall scor

revealed residents had performed significantly more tonsillectomies in
technique group (47.8) compared to the electrosurgical technique grou

There was no difference in number of procedures performed, mean TB
of patient between complex and standard cases. The number of tonsil p
mean TBC score significantly affected the likelihood of a GRS rating
Table 1, (P<0.001 for both). The use of radiofrequency technique was
with competency (P<0.001). However, patient age did not significantl
likelihood of competency.

The adjusted odds ratios to predict competence by procedure number,
mean Tonsil TBC score can be found in **Table 2**. Each tonsillectomy p
increased the likelihood of a GRS rating of competency by 6.3% (P=0
1.110). Moreover, with a 1.0 point increase in mean score on the tonsi
rating of competency increased by a factor 2.71 (P=0.006, 95% CI of
adjusting for the number of procedures performed and the mean TBC
technique was no longer associated with the likelihood of a rating of c
CI of 0.610 to 11.488).

Using the procedure number to predict competency, there is an 80% li
rating of competence at 31 standard tonsillectomies and a 95% likeliho
tonsillectomies [**Figure 2**]. Using the mean score on the tonsil TBC, th
of a subjective rating of competence at a score of 3.99 and a 95% likel
[**Figure 3**]. The receiver operator characteristic (ROC) curve for this m
under the curve (AUC) of 0.9220 [**Figure 4**].

To assess construct validity (i.e, increased procedure numbers correlat
likelihood of competency on the GRS) we compared the mean tonsil T
likelihood of competency for residents performing ≤ 10 cases versus th
tonsillectomy cases. For those who had performed less than 10 proced

(K=0.64, *P*=0.02), 1 had moderate agreement (K=0.54, *P*=0.08), 2 had
P=0.01, K=0.28, *P*=0.19) and 1 had slight agreement (K=0.16, *P*=0.35
competency also had perfect agreement (K=1, *P*=0.01) [**Table 4**].

The mean time for completion of the survey was 87±66 seconds. A su
reported here) found the instrument to be easy to understand, compreh

Discussion:

As hypothesized, residents who had performed more tonsillectomies w
assessed as competent. Residents deemed competent had also perform
tonsillectomies prior to assessment than those rated not competent and
tonsil TBC correlated with a greater likelihood of being rated compete
findings, each tonsillectomy performed resulted in a 6.3% increase in
deemed competent and each 1.0 increase in the mean tonsil TBC score
of competence by a factor 2.7. We found that there was a 95% likeliho
residents had performed 48 tonsillectomies or had a tonsil TBC score
reliability was perfect to moderate for 7 of the 10 TBC procedure step
5 residents.

These results are similar to previous OSATS studies which have valida
for the assessment of resident surgical skills.[4-14] TBC and GRS have b
of simulation models for otolaryngology surgical skills which have sim
construct validity and high interrater reliability. These include OSATS
management[4], endoscopic sinus surgery[5,10,11], mastoidectomy[6, 12], pedi
[13], and transoral robotic surgery training[9]. Similarly OSATS are now
operating room with a previous pilot study demonstrating feasibility a
tonsillectomy specific TBC and GRS[14]. The tonsil OSATS demonstra
reliability, which is comparable to reported findings for other OSATS
flexible gastrointestinal endoscopy skills[15], which was 0.95, and wit
surgery, which was 0.998[16].

Contemporary medical education, first introduced by Flexner, Halsted foundation of surgical education. The main tenants of surgical educati responsibility, supervision, and mentorship; these factors remain integ surgeons. However, as other professions move toward competency ass has lagged behind. In light of these factors, societal and professional p surgical education toward objective measures of competency[1] and obj surgical competency will likely be required in the future by regulatory authorities.

Additional challenges not envisioned by Flexner, Halsted and Osler in resident work-hour restrictions in the United States in 2003. Work hou restricted in 2011 to limit the number of consecutive hours that reside hour restrictions are particularly challenging for surgical educators as surgical case exposure in order to achieve competency in a wide array surgical procedure, it is assumed that the total number performed as re with competency. However, the number needed to achieve competenc procedure and by surgeon.

We hypothesized that understanding the learning curve for a procedur objectively assess surgical skills would allow surgical trainees and the optimally tailor resident's operative experience. These regular assessm allowed for regular feedback in order to facilitate more efficient learni competence is achieved, the resident could spend time working to get status, and allow residents and their instructors to apportion their time achieve competence in all the required procedures.

Our data suggests that after 48 tonsillectomies, there is a 95% likeliho within the current mean of 114 tonsillectomies performed by residents performance of tonsillectomies beyond those needed to achieve compe opportunity for residents to achieve proficiency or expert status, but gi surgical educators may use this information to decide that this time mi residents learning additional procedures. Tonsil OSATS can be used t

used by residents of any PGY level. Secondly, only five cases were re
reliability; this data is limited (N=5) due to the difficulty obtaining a s
otolaryngologist who had no knowledge of the resident, PGY, or case
results demonstrated a high level of agreement between evaluators, wh
interrater reliability findings for other specialty OSATS[15,16]. In additio
number of cases, 7 of 10 tasks on the Tonsil PCS had moderate to per
Despite these limitations, we found the tonsil OSATS to be feasible an
validity. Future research is warranted across a larger sample size to co
interrater reliability.

Conclusion:

Tonsil OSATS provide a valid, and feasible measure of otolaryngolog
performing tonsillectomy as well as providing a tool for formative fee
conjunction with the judgement of the supervising surgeon. Tools such
as regulatory agencies and credentialing authorities require objective c
competence.

REFERENCES

1. Accreditation Council for Graduate Medical Education.

ACGME program requirements for graduate medical education in otol

2. Brown DJ, Thompson RE, Bhatti NI. Assessment of operative comp

residency: Survey of US program directors. *Laryngoscope*. 2008;118(

10.1097/MLG.0b013e31817e2c62; 10.1097/MLG.0b013e31817e2c62

3. Reznick R, Regehr G, MacRae H, Martin J, McCulloch W. Testing

innovative “bench station” examination. Am J Surg. 1997. Mar;173(3)

4. Goff B, Mandel L, Lentz G, Vanblaricom A, Oelschlager AM, Lee

Nielsen P. Assessment of resident surgical skills: is testing feasible? A

Apr;192(4):133-8.

5. van Hove PD, Tuijthof GJ, Verdaasdonk EG, Stassen LP, Dankelma

of technical surgical skills. *Br J Surg*. 2010;97(7):972-987. doi: 10.100

10.1002/bjs.7115.

7. Laeeq K, Infusino S, Lin SY, et al. Video-based assessment of oper
endoscopic sinus surgery. *Am J Rhinol Allergy*. 2010;24(3):234-237. d
10.2500/ajra.2010.24.3434; 10.2500/ajra.2010.24.3434.

8. Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment to
mastoidectomy. *Laryngoscope*. 2009;119(12):2402-2410. doi: 10.1002
10.1002/lary.20678.

9. Swift SE, Carter JF. Institution and validation of an observed struct
technical skills (OSATS) for obstetrics and gynecology residents and
Gynecol. 2006;195(2):617-21; discussion 621-3. doi: 10.1016/j.ajog.2

10. Jabbour N, Reihsen T, Payne NR, Finkelstein M, Sweet RM, Sidm
assessment tools for pediatric airway endoscopy simulation. *Otolaryng
Surgery*. 2012 Dec;147(6):1131-5.

11. Sobel RH, Blanco R, Ha PK, Califano JA, Kumar R, Richmon JD.
comprehensive competency-based transoral robotic surgery training cu
dissection models. *Head and Neck*. 2016. Oct;38(10):1553-63.

13. Laeeq K, Infusino S, Lin SY, et al. Video-based assessment of ope
endoscopic sinus surgery. *Am J Rhinol Allergy*. 2010;24(3):234-237. d
10.2500/ajra.2010.24.3434; 10.2500/ajra.2010.24.3434.

14. Laeeq K, Bhatti NI, Carey JP, et al. Pilot testing of an assessment t
mastoidectomy. *Laryngoscope*. 2009;119(12):2402-2410. doi: 10.1002
10.1002/lary.20678.

15. Ishman SL, Brown DJ, Boss EF, et al. Development and pilot testi
competency assessment tool for pediatric direct laryngoscopy and rigi
Laryngoscope. 2010;120(11):2294-2300. doi: 10.1002/lary.21067.

16. Ahmed A, Ishman SL, Laeeq K, Bhatti NI. Assessment of improve
skills in the operating room for tonsillectomy. *Laryngoscope*. 2013. do
10.1002/lary.24023.

17. Accreditation Council for Graduate Medical Education.
NATIONAL RESIDENT REPORT (comprehensive and key indicator

Tables

Table 1: Predictors of Tonsillectomy Competency

Procedure *(number of tonsillectomies performed).* ***Tonsil TBC*** *(tonsil checklist).* ***Age*** *(age of patient in years).* ***Competent*** *(overall rating of tonsillectomy independently).* ***Non-competent*** *(overall rating of not co tonsillectomy independently).*

Table 2: Adjusted odds ratios of the overall subjective rating of co evaluated using the tonsillectomy objective structured assessment (OSATS), n=163 (4 techniques were missing). For technique, 117 us radiofrequency and 5 were classified as other. The model included the TBC score due to their significance while the technique was included

Procedure *(number of tonsillectomies performed).* ***Tonsil TBC*** *(Tonsi checklist mean score).*

For procedure and tonsil TBC, odd ratios represents a 1 unit increase

Table 3: Mean tonsil task-based checklist (TBC) score and global competency rating for residents who had previously performed ≤ those who had performed ≥50 tonsillectomies.

Table 4: Interrater Reliability for each item included in the Tonsil checklist (TBC) and competency evaluation (n=5)

Tonsil TBC *(Tonsillectomy task-based checklist mean score).*

Figure Legend

Figure 1: Tonsillectomy objective structured assessment of technical
to assess residents to determine if they were competent on the procedu

Figure 2: Likelihood of competency by procedure number for residen
tonsillectomy.

Procedure *(number of previous tonsillectomies performed).*

Figure 3: Likelihood of competency by mean score of tonsil task-base
residents performing tonsillectomy.

Tonsil TBC *(tonsillectomy task based checklist mean score.)*

Figure 4: Receiver Operating Characteristic (ROC) curve for the tons
assessments of technical skills (OSATS).

Tonsillectomy OSATS Task Specific Rating Assessment

A faculty member is to complete this from after observing a resident perform a tonsillectomy. Participation is voluntary. By completing this form you are implicitly consenting to participate. Any reported or published data from this study will be presented without any identifying markers.

Faculty name:_____ Date:_____

Resident ID: _____ PGY Level: _____

Number of tonsillectomies performed: _____

Age of patient (years) _____

Technique:

- Electrosurgical (i.e. Bovie)
- Suction electrosurgical (i.e. suction Bovie)
- Cold steel
- Colbation-Total tonsillectomy
- Colblation-Intracapsular
- Microdebrider-Total tonsillectomy
- Microdebrider- Intracapsular
- Other _____

	Verbal instruction and demonstration	Requires instruction with errors	Independent with errors	Independent without errors	Independent and efficient	Not performed
1. Patient positioning and draping						NA
2. Atraumatic mouth gag placement						NA
3. Gag suspension						NA
4. Adequate exposure of the tonsils						NA
5. Grasps tonsils appropriately						NA
6. Finds the plane						NA
7. Dissects in the correct plane						NA
8. Obtains hemostasis						NA
9. Suctions stomach						NA
10. Removes gag safely while respecting ETT placement						NA

Attending rating of case complexity:

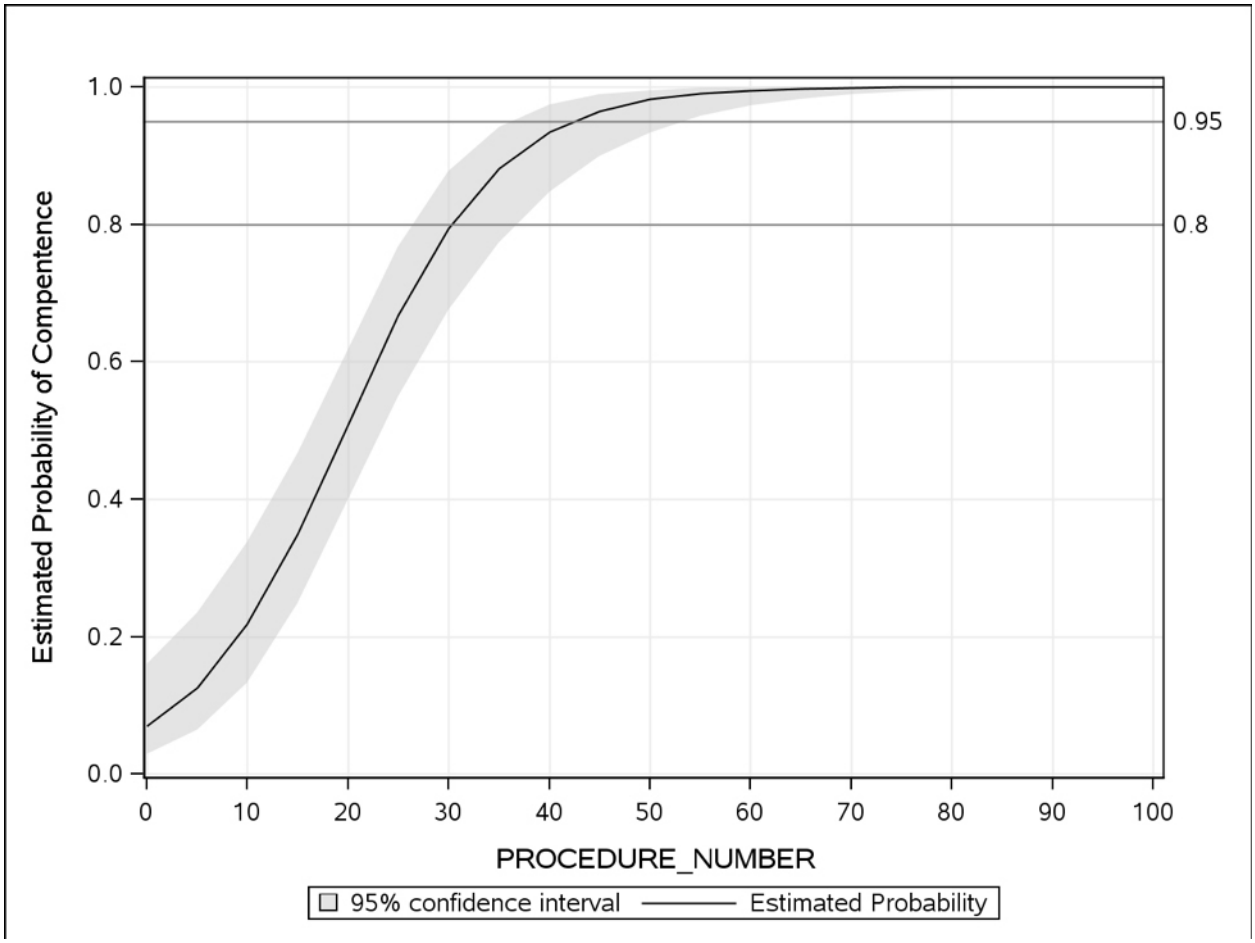
- ☐ Standard case that most general otolaryngologists would feel comfortable performing
- ☐ Challenging case for most general otolaryngologists

This was a challenging case because_____

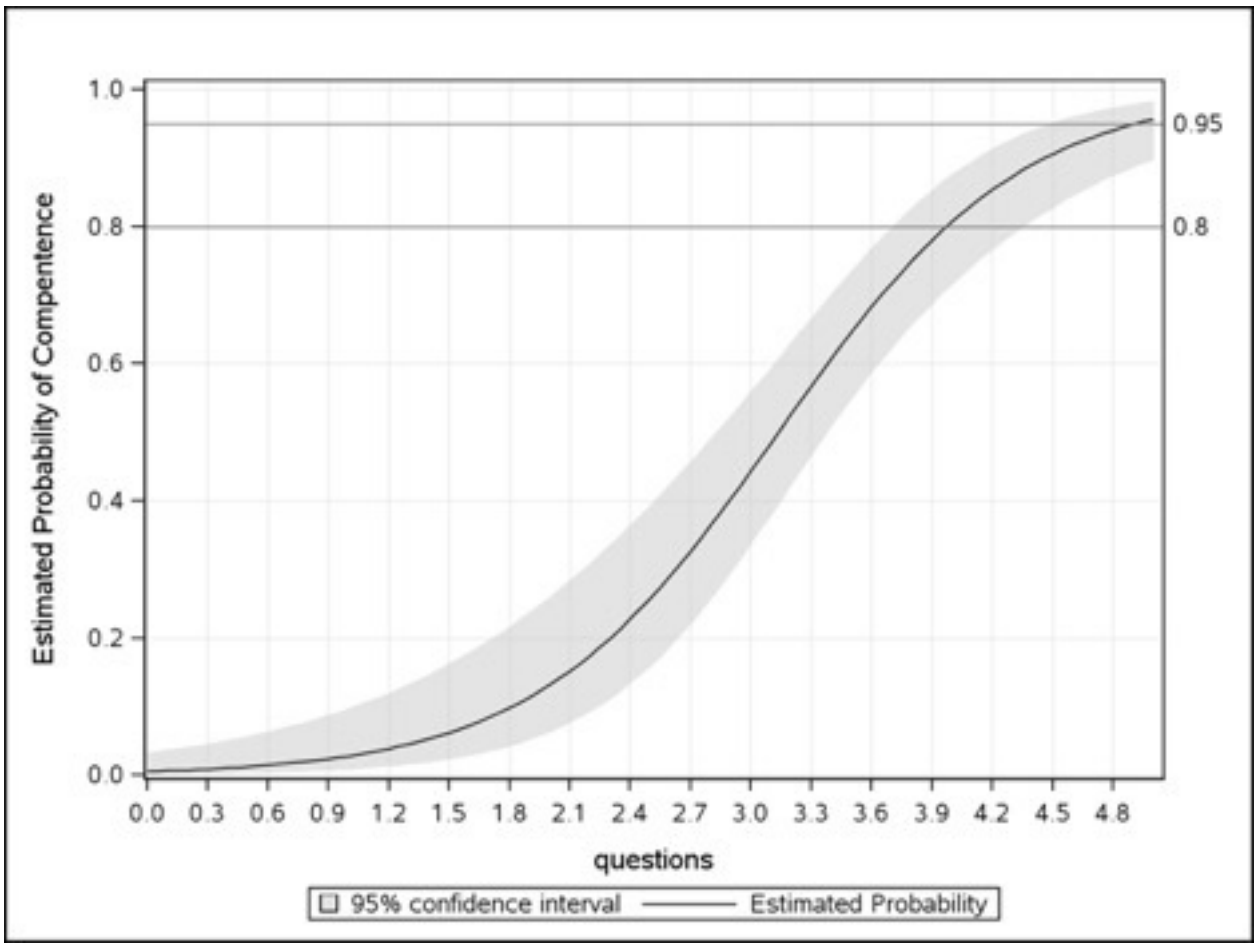
Based on the overall performance on this case, is the resident capable of independently performing a tonsillectomy in a safe and competent manner?

Attending Response: Yes No

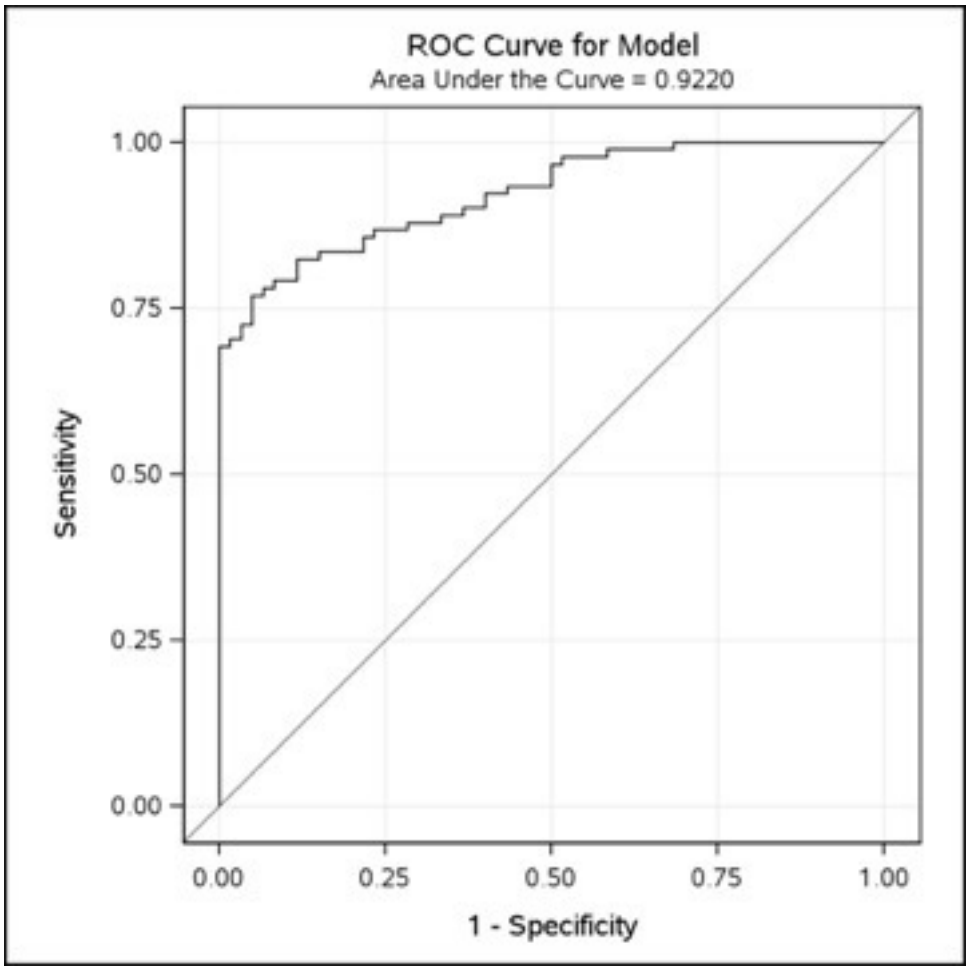
LARY_28739_Figure-1.jpg



LARY_28739_Figure-2.jpg



LARY_28739_Figure-3.jpg



LARY_28739_Figure-4.jpg

Tables

Table 1: Predictors of Tonsillectomy Competency

			Global Rating Scale (
Variables	Total *N=167*	Non-competent N=68	Comp N=
Procedure			
N	167	68	9
Mean ± SD	31.8 ± 32.2	13.5 ± 11.6	44.4 ±
Median (min - max)	24.0 (0.0 - 300.0)	10.5 (0.0 - 60.0)	39.0 (300
Tonsil TBC			
N	167	68	9
Mean ± SD	3.5 ± 1.1	2.6 ± 1.0	4.0 ±
Median (min - max)	3.7 (0.8 - 5.0)	2.8 (0.8 - 4.3)	4.2 (1.3
Age of patient			
N	162	65	9
Mean ± SD	5.7 ± 3.5	5.8 ± 3.8	5.6 ±
Median (min - max)	5(1-17)	5(1-16)	5(1-
Technique			
N	163	67	9
Electrosurgical	117 (71.8%)	56 (83.6%)	61 (63

Table 2: Adjusted odds ratios of the overall subjective rating of co evaluated using the tonsillectomy objective structured assessment (OSATS), n=163 (4 techniques were missing). For technique, 117 us radiofrequency and 5 were classified as other. The model included the TBC score due to their significance while the technique was included

	Adjusted Odds Ratio	Co
Procedure Number	1.063	1
Tonsil TBC	2.71	1
Technique (Radiofrequency vs Others)	2.647	0

Procedure *(number of tonsillectomies performed).* ***Tonsil TBC*** *(Tonsi checklist mean score).*

For procedure and tonsil TBC, odd ratios represents a 1 unit increase

Table 3: Mean tonsil task-based checklist (TBC) score and global competency rating for residents who had previously performed ≤ those who had performed ≥50 tonsillectomies.

	Procedure Number		
	<10 N=34	*>50 N=29*	*P Value*
Tonsil TBC			*<.001*[G]
N	34	29	
Mean ± SD	2.0 ± 0.8	4.3 ± 0.5	
Median (min - max)	2.0 (0.8 - 3.5)	4.5 (3.2 - 5.0)	
Competence			*<.001* [C+]
No	30 (88.2%)	2 (6.9%)	
Yes	4 (11.8%)	27 (93.1%)	

[C+]Chi-square test with Yates's continuity correction; [G]GEE

Procedure *(number of tonsillectomies performed).* ***Tonsil TBC*** *(Task score).*

Table 4: Interrater Reliability for each item included in the Tonsil checklist (TBC) and competency evaluation (n=5)

TBC	Kappa Coefficient
Patient positioning and draping	0.642
Atraumatic mouth gag placement	1.000
Gag suspension	1.000
Adequate exposure of the tonsils	0.375
Grasps tonsils appropriately	1.000
Finds the plane	0.642
Dissects in the correct plane	0.285
Obtains hemostasis	0.166
Suctions stomach	0.545
Removes gag safely while respecting ETT placement	1.000
Overall Rating of competency	1.000