



Assessment of the Reliability of the Fiberoptic Endoscopic Evaluation of Swallowing as an Outcome Measure in Patients Undergoing Revision Anterior Cervical Discectomy and Fusion

Matthew S. Erwood¹, Beverly C. Walters¹, Timothy M. Connolly⁴, William R. Carroll², Duane V. Trahan³, Abby W. Graves³, Bonita S. Agee¹, Mark N. Hadley¹

■ **BACKGROUND:** Dysphagia is one of the most common complications of anterior cervical spine surgery, and there is a need to establish that the means of testing for it are reliable and valid. The objective of this study was to measure observer variability of the fiberoptic endoscopic evaluation of swallowing (FEES) test, specifically when used for evaluation of dysphagia in patients undergoing revisionary anterior cervical decompression and fusion (ACDF).

■ **METHODS:** Images from patients undergoing revision ACDF at a single institution were collected from May 1, 2010, through July 1, 2014. Two senior certified speech pathologists independently evaluated the swallowing function of patients preoperatively and at 2 weeks postoperatively. Their numeric evaluations of the Rosenbeck Penetration-Aspiration Scale and the Swallowing Performance Scale during the FEES were then compared for interrater reliability.

■ **RESULTS:** Positive agreement between raters was 94% for the preoperative Penetration-Aspiration Scale (prevalence-adjusted bias-adjusted κ , 0.77). The postoperative Penetration-Aspiration Scale showed reliability coefficients for κ , Kendall's W, and intraclass correlation coefficient (ICC) of 0.34 (fair agreement), 0.70 (extremely strong agreement), and 0.35 (poor agreement), respectively. The preoperative Swallowing Performance Scale showed strong agreement, with a Kendall's W coefficient of 0.68, and fair reliability, with an ICC of 0.40. The postoperative

Swallowing Performance Scale indicated extremely strong agreement between raters, with a Kendall's W of 0.82, and good agreement, with an ICC of 0.53.

■ **CONCLUSIONS:** The FEES test appears to be a reliable assessor of dysphagia in patients undergoing ACDF and may be a useful measure for exploring outcomes in this population.

INTRODUCTION

Dysphagia and resultant aspiration or malnutrition can adversely affect outcomes in individuals with brain and spinal pathology.¹ In particular, dysphagia is one of the most common complications of anterior cervical decompression and fusion surgery^{2,3} and can result from numerous mechanisms, including soft tissue inflammation and edema and damage to the recurrent or superior laryngeal nerves. Injury to the recurrent laryngeal nerve can manifest as dysphonia or dysphagia, with an incidence of up to 11% in the current literature.²⁻⁴ The rate of recurrent laryngeal nerve injury is higher in revision anterior cervical surgeries, at approximately 14%.⁵

In today's practice landscape, reimbursement strategies from payers have increasingly focused on patient satisfaction postoperatively, an outcome influenced heavily by the occurrence of complications and/or adverse events. Consequently, it is becoming increasingly important for neurosurgeons to identify and track their most common surgical complications.⁶ As we measure outcomes in anterior cervical spine surgery, means of testing

Key words

- Dysphagia
- Fiberoptic endoscopic evaluation of swallowing
- Reliability
- Reoperative ACDF surgery

Abbreviations and Acronyms

- ACDF:** Anterior cervical discectomy and fusion
FEES: Fiberoptic endoscopic evaluation of swallowing
ICC: Intraclass correlation coefficient
IRR: Interrater reliability
MDADI: MD Anderson Dysphagia Inventory
PABAK: Prevalence-adjusted bias-adjusted κ

From the ¹Department of Neurosurgery, ²Division of Otolaryngology, Department of Surgery, and ³Department of Speech Pathology, University of Alabama at Birmingham, Birmingham, Alabama, USA; and ⁴Department of Surgery, University Hospital Geelong, Geelong, Victoria, Australia

To whom correspondence should be addressed: Matthew S. Erwood, M.D.
 [E-mail: merwood@uabmc.edu]

Citation: *World Neurosurg.* (2019) 130:e199-e205.
<https://doi.org/10.1016/j.wneu.2019.06.028>

Journal homepage: www.journals.elsevier.com/world-neurosurgery

Available online: www.sciencedirect.com

1878-8750/\$ - see front matter Published by Elsevier Inc.

and evaluation should be reliable and validated. Given that dysphagia is one of the most common complications of anterior cervical spine surgery, particularly anterior cervical decompression and fusion (ACDF), the need for reliable and valid swallowing evaluation is essential.⁷

Two assessments—fiberoptic endoscopic evaluation of swallowing (FEES) and the modified barium swallow study—are considered the gold standards for assessing swallowing function.⁸ In a previous study, FEES was the outcome measure used to determine swallowing function among patients undergoing revision ACDF surgery.⁵ The FEES test involves direct visualization of a patient's vocal cords and pharynx while swallowing various substances through the use of a flexible laryngoscope⁹ (Figures 1–4). This visualization is then documented using 2 assessment scales that categorize different aspects of function: the Swallowing Performance Scale and the Penetration-Aspiration Scale (Figure 5).^{10,11} Although the FEES test has been shown to be an effective tool to evaluate the presence or absence of swallowing disorders in other situations, to our knowledge there are little published data to establish its interrater reliability (IRR) in patients undergoing anterior cervical spine procedures.^{12–15} Consequently, the objective of this study was to measure observer variability of FEES, specifically when used for evaluating dysphagia in patients undergoing revisionary ACDF.

METHODS

FEES studies from patients undergoing revision anterior cervical discectomy and fusion procedures at a single institution were enrolled from May 1, 2010, through July 1, 2014. Institutional Review Board approval was obtained before enrollment, and all participating patients consented to inclusion in the study. A total of 55 patient studies were used in the study; several of the patients had undergone previous anterior cervical spine surgery at 3 or more surgical levels.

All patients enrolled in the study underwent evaluation by a single senior certified speech and language therapist. There was an initial preoperative evaluation, with an additional follow-up evaluation at 2 weeks postoperatively. Patients with documented vocal cord or swallowing dysfunction were evaluated further at 3 months and 6 months postoperatively, if necessary. Data on dysphagia and vocal cord paralysis due to injury to the recurrent laryngeal nerve were recorded by the speech therapist. These were determined through the FEES examination and video-laryngostroboscopy, respectively. For the FEES, patients were administered thin liquids, thickened liquids, and solid food, and swallowing function was assessed according to the Swallowing Performance Scale and the Penetration-Aspiration Scale.^{10,11} The initial studies were stored electronically, allowing repeat analysis at a later date.

For this reliability study, a second senior certified speech pathologist independently evaluated archived videos of the swallow functions of individual patients, mimicking normal clinical practice.¹⁶ The observations of the speech pathologists were recorded using the numeric evaluations of the Penetration-Aspiration Scale and the Swallowing Performance Scale and were then compared between the 2 raters to assess IRR. An example of the data sheet used in the study is shown in Figure 5.

Outcomes for the study were varying degrees of stasis, penetration of the vocal cords, or aspiration as seen during the FEES test. Exposure for patients was the performance of a reoperative anterior cervical discectomy and fusion. Predictors of dysphagia were retraction and stretch injury during performance of the ACDF, complicated by the scarring inherent with a revisionary procedure. The presence of swallowing dysfunction was also an indicator of potential damage to the recurrent laryngeal nerve.¹⁷ The speech pathologists were blinded to each other's evaluations to protect against introduction of potential bias.

This study was designed with special attention to several critical factors for reliability studies that have previously been outlined in the literature. Requirements included the identification of an appropriate research question and the use of both raters and patients to provide a true sample of the individuals who would use this test in a clinical setting. The Penetration-Aspiration Scale and Swallowing Performance Scale were chosen as the means of comparison because these numerical interpretations of the FEES test are used commonly in practice to define swallowing abnormalities. Appropriate statistical measures were used to interpret the data, as described below. Because this was a cross-sectional consecutive sample, there were no quantitative variables to analyze.¹⁶

The estimate for sample size for the number of ratings in this study was based on $\alpha = 0.05$, 80% power ($\beta = 0.20$), and a minimally acceptable reliability estimate of 0.60, because this is often considered a cutoff for moderate reliability. Using the method described by Walter et al.¹⁸ with 2 raters and looking at the difference between an anticipated reliability coefficient (ρ_1) of 0.80 and a minimally acceptable reliability coefficient (ρ_0) of 0.60, a sample of size of at least 39 comparisons was indicated. This would provide 80% power for testing $H_0: \rho_1 \leq \rho_0$ versus $H_1: \rho_1 \geq \rho_0$ using the foregoing specifications.

Statistical Analysis

IRR for the FEES parameters was calculated using weighted κ with Fleiss–Cohen weights, Kendall's coefficient of concordance (W), and intraclass correlation coefficient (ICC). Landis and Koch's categorization¹⁹ interprets κ values of 0.81–1.00 as almost perfect, 0.61–0.80 as substantial agreement, 0.41–0.60 as moderate agreement, 0.21–0.40 as fair agreement, and 0.00–0.20 as poor to slight agreement.

Another statistic, the prevalence-adjusted bias-adjusted κ (PABAK), adjusts the κ for imbalances caused by differences in the prevalence and presence of bias.²⁰ This additional indicator measures observer reliability in instances where the high prevalence of a rating has resulted in low variability, as in the case of high agreement, and thus low IRR. Kendall's W is a test of concordance specific for ordinal categories and quantifies agreement of 2 or more raters. A Kendall's W coefficient of concordance of 0.71–9.0 is considered to indicate extremely strong agreement, 0.51–0.70 to indicate strong agreement, 0.31–5.0 to indicate moderate agreement, and 0.11–0.30 to indicate weak agreement.²¹ Both κ and Kendall's W were computed using the SAS MAGREE macro version 1.0 (SAS Institute, Cary, North Carolina, USA).²²

IRR coefficients were also generated using the ICC, a popular method for quantifying the extent of agreement among raters by



Figure 1. Appearance of normal vocal cords on the fiberoptic endoscopic evaluation of swallowing.



Figure 3. Abnormal swallowing study—stasis.

comparing the variability of raters rating the same subject to the total variation across all ratings and subjects. The ICC values presented here are based on ICC model 2 of Shrout and Fleiss,²³ a 2-way analysis of variance random-effects model with absolute agreement (ICC_{2,1}). Model 2 was chosen because each rater rated the entire sample of subjects, and the 2 raters who gave the clinical assessments were believed to represent a random sample from a larger population of senior speech pathologists. In this way, this model permits generalizability to other potential raters.²³ Agreement measured by the ICC is interpreted as excellent at >0.75, fair to good at 0.40–0.75, and poor at <0.40.²⁴

All statistical analyses were performed using SAS version 9.4 (SAS Institute, Cary, North Carolina, USA), with the exception of the ICCs, which were computed using SPSS version 24.0 (IBM,

Armonk, New York, USA). A *P* value <0.05 was considered to indicate statistical significance.

RESULTS

Ninety-eight total FEES examinations (54 preoperative and 44 postoperative) from 55 patients were chosen for the study, providing sufficient power. The study cohort included 26 males (47%) and 29 females (53%), with an average age of 53 years. The most common reason for a reoperative ACDF procedure was radiculopathy (74.5%), with myelopathy, instability, and pseudoarthrosis as additional indications. Twenty-nine patients (53%) underwent fusion of 3 or more cervical vertebrae in at least 1 preceding operation. All 55 patients were treated via a right-sided approach.¹⁷ This patient population thus represented the typical



Figure 2. Normal phonation during the fiberoptic endoscopic evaluation of swallowing.



Figure 4. Abnormal swallowing study—penetration.

Flexible/Fiberoptic Endoscopic Evaluation of Swallowing (FEES)

Date of Surgery: _____ Current Diet: _____

Consistencies Administered: ___ Pureed ___ Regular ___ Thin liquid ___ Thickened Liquid

Indirect Oral Phase Assessment

Oral Phase Swallowing Assessment was normal / abnormal

Direct Pharyngeal Phase Assessment

Signs: (Abnormal)

- ___ Presence of residual secretions prior to PO intake
- ___ Erythematic/Edematous Arytenoid Mucosa
- ___ Initiation of Pharyngeal Swallow Reflex (norm≈1-2 sec)
- ___ Stasis in Valleculae
- ___ Stasis in Left/Right Channels
- ___ Stasis in Left/Right Pyriform Sinus
- ___ Stasis on Posterior/Left/Right Pharyngeal walls
- ___ Reduced Laryngeal Closure
 - ___ Left ___ Right
- ___ Piecemeal Deglutition

Interpretation:

- ___ Swallow Reflex Delay
- ___ Clinically Absent Swallow Reflex <30 sec
- ___ Impaired Tongue Base Retraction
- ___ Impaired Pharyngeal Constriction
 - ___ Left ___ Right
- ___ Impaired Hyolaryngeal Excursion
- ___ Impaired Airway Protection
- ___ Impaired Sequential Timing
- ___ Questionable Esophageal Reflux
- ___ **Within Normal Limits**

Laryngeal Penetration Yes / No Sensitive / Nonsensitive **Aspiration** Yes / No Sensitive / Nonsensitive
 Consistency: _____ Consistency: _____

Impressions/Ratings

Swallowing Performance Scale

- 1- Normal
- 2- WFL –abnormal oral or pharyngeal stage but able to eat regular diet without modifications or swallowing precautions.
- 2- Mild impairment –mild dysfunction in oral or pharyngeal stage. Requires modified diet or therapeutic swallowing precautions.
- 3- Mild-Moderate impairment –need for therapeutic precautions – mild dysfunction in oral or pharyngeal stage, requires modified diet and swallowing precautions to minimize risk of aspiration.
- 4- Moderate impairment –moderate dysfunction in oral or pharyngeal stage, aspiration noted on exam, requires modified diet and swallowing precautions to minimize risk of aspiration.
- 5- Moderate-severe dysfunction –requires supplemental central feeding support – moderate dysfunction in oral or pharyngeal stage. Aspiration noted on exam, requires modified diet and swallowing precaution to minimize risk of aspiration, needs enteral feeding support.
- 6- Severe impairment –severe dysfunction with significant aspiration or inadequate oropharyngeal transit to esophagus, NPO, requires primary enteral feeding support.

Rosenbeck Aspiration Penetration Scale

- 1- Material does not enter airway.
- 2- Material enters the airway, remains above the level of the vocal folds, and is ejected from airway.
- 3- Material enters the airway, remains above the level of the vocal folds, and is not ejected from airway.
- 4- Material enters the airway, contacts the vocal folds, and is ejected from the airway.
- 5- Material enters the airway, contacts the vocal folds, and is not ejected from the airway.
- 6- Material enters the airway, passes below the vocal folds, and is ejected into the larynx or out of the airway.
- 7- Material enters the airway, passes below the vocal folds, and is not ejected into the larynx or out of airway.
- 8- Material enters the airway, passes below the vocal folds, and no effort is made to eject.

Figure 5. Example assessment used for the fiberoptic endoscopic evaluation of swallowing, including the

Swallowing Performance Scale and the Penetration-Aspiration Scale.

Table 1. Statistical Methods for Diagnostic Agreement

Method	Agreement	Coefficient Values
Weighted κ ¹⁹ with Fleiss-Cohen (quadratic) weights	Almost perfect	0.81–1.00
	Substantial	0.61–0.80
	Moderate	0.41–0.60
	Fair	0.21–0.40
	Poor	0.00–0.20
Kendall's coefficient of concordance (W) ²¹	Extremely strong	0.71–0.90
	Strong	0.51–0.70
	Moderate	0.31–0.50
	Weak	0.11–0.30
Intraclass correlation coefficient (ICC) ²⁴	Excellent	0.75–1.00
	Good	0.60–0.74
	Fair	0.40–0.59
	Poor	0.00–0.40

scope of patients in whom the FEES test would be applied in practice.

Reliability data for the FEES test as assessed using the Penetration-Aspiration Scale and the Swallowing Performance Scale and compared using κ , Kendall's W coefficient of concordance, and the ICC are provided in **Tables 1** and **2**.

Positive agreement between raters was 94% for the preoperative Penetration-Aspiration Scale. The lack of variability in the range of scores on the preoperative assessments (i.e., most patients with normal or close to normal function) essentially excluded those results from evaluation by κ and ICC. However, the additional κ statistic described above, PABAK, was calculated as 0.77 for the preoperative Penetration-Aspiration Scale, consistent with substantial agreement between raters.²⁵

On the postoperative Penetration-Aspiration Scale, reliability coefficients for κ , Kendall's W, and ICC were 0.34 (fair agreement),

0.70 (extremely strong agreement), and 0.35 (poor agreement), respectively. As discussed below, the ICC and κ results were more strongly affected than Kendall's W results by the prevalence and variability of the data.²⁶

Swallowing Performance Scale findings were better than those for the Penetration-Aspiration Scale. Preoperative assessments showed strong agreement, with a Kendall's W coefficient of 0.68, and fair reliability, with an ICC of 0.40. Meanwhile, postoperative assessment indicated extremely strong agreement between raters, with a Kendall's W of 0.82 and an ICC of 0.53 interpreted as showing good agreement.

DISCUSSION

The FEES test is widely considered the gold standard for evaluating dysphagia. In principle, a gold standard should prove reliable and valid across evaluations and patient populations. Therefore, the FEES test not only should provide accuracy in determining the diagnosis of dysphagia, but it should do so consistently across users of the test.^{27,28} Because neurosurgeons frequently assess their own outcomes using such methods as the FEES test, it is important that they understand not only the outcome, but also how objective and robust the assessment is, as measured by its reliability.

Despite the lack of previously reported data showing the reliability of the FEES test in ACDP-related dysphagia, this test is commonly used as an objective measure of swallowing assessment in this target population.^{4,5} As shown in **Figure 5**, 2 numerical scales are included as part of an FEES evaluation: the Swallowing Performance Scale and the Penetration-Aspiration Scale. Users of the FEES laryngoscope, often speech pathologists, assign a numerical score based on the level of dysfunction identified in the patient's ability to swallow.^{15,29} Explanations for each numerical score are included in **Figure 5** as well. The reliability of the FEES test, as measured by the Swallowing Performance Scale and the Penetration-Aspiration Scale, was the focus of this study.

Subjective, patient-reported assessments of swallowing exist as well. Chief among these are questionnaire-based assessments of swallowing function, such as the MD Anderson Dysphagia Inventory (MDADI). The MDADI scale uses a series of swallowing questions across multiple subscales, allowing the patient to rate his or her own swallowing capability and its effect on emotional

Table 2. FEES Test IRR Estimates

Parameter	κ	SE	P Value	Kendall's W	P Value	ICC	P Value
Penetration-Aspiration Scale, preoperative	0.77*						
Penetration-Aspiration Scale, postoperative	0.34	0.19	0.001	0.70	0.01	0.35	0.003
Swallowing Performance Scale, preoperative	0.39	0.14	0.0002	0.68	0.004	0.40	<0.0001
Swallowing Performance Scale, postoperative	0.52	0.08	<0.0001	0.82	<0.0001	0.53	<0.0001

Significant P values (<0.05) are in bold type.
 FEES, fiberoptic endoscopic evaluation of swallowing; IRR, interrater reliability; SE, standard error; ICC, intraclass correlation.
 *Prevalence-adjusted, bias-adjusted κ (PABAK).

status and quality of life. The questions are then assigned a numerical score and tabulated to form an overall assessment.³⁰

Detailed FEES data in patients with ACDF undergoing reoperation have been reported recently.¹⁷ A recent study examined patients undergoing revision anterior cervical spine surgery with preoperative and postoperative FEES and MDADI assessments. The postoperative evaluations were obtained at 2 weeks after revision ACDF. The authors found that 30 of 67 patients (44.8%) exhibited abnormalities on the postoperative FEES test. The majority of these swallowing abnormalities were mild, although 8 patients required swallowing precautions in the postoperative period and 2 patients required a dietary change. In comparison, roughly 60% of patients reported a subjective decline in swallowing function on the MDADI, as opposed to the roughly 45% detected by FEES.¹⁷

Numerous studies in the literature support the effectiveness of the FEES test as a predictive tool. It can be used to assess the severity of disease in amyotrophic lateral sclerosis and has proven to be an effective tool for determining the timing of patient dietary interventions.^{31,32} In 1 such study, aspiration pneumonia did not occur in any of the 22 patients who began oral feeding based on their FEES results.³² Moreover, the FEES test has proven effective and safe for use across age groups, including the pediatric population, to evaluate general swallowing disorders.^{33,34}

Although not been previously studied in the clinical scenario of ACDF surgery, the FEES test has been shown to be both valid and reliable in other situations, especially when used with the Penetration-Aspiration Scale. This has been true for both intrarater reliability and IRR, regardless of the interval between testing, and for multiple swallowing pathologies.^{12,15,35} One recent study found an IRR of 0.763 (Kendall's W) for white light FEES dysphagia evaluations using the Penetration-Aspiration Scale.¹⁴ Whereas the reported correlation coefficients for reliability vary widely in the literature, ranging as high as 0.97 when used by advanced clinicians, this scale has consistently proven to be reliable.¹² Although data on the reliability and validity of the Swallowing Performance Scale are less prevalent, this scale remains a frequently used tool for assessing swallowing function and is commonly used to rate the severity of dysphagia in patients with head and neck cancer undergoing chemotherapy and radiation treatment.^{11,29}

To our knowledge, the present study is the first to examine the reliability of the FEES test for dysphagia in patients undergoing anterior cervical spine reoperation. Overall, our results indicate that the FEES test is a reliable assessment for this outcome. The Penetration-Aspiration Scale results were plagued by the very high

prevalence of scores of 1 (normal) in the raters' assessments to the exclusion of other ratings, resulting in low variability among ratings and subsequently low IRR. This phenomenon related to the κ statistic is known as Cicchetti's paradox.³⁶ To counteract this, the PABAK was calculated, and the adjusted κ confirmed a high IRR in the preoperative population. Because the scoring did not cover the full range of values being tested (because most patients had completely normal function), the kappa, Kendall's W, or ICC calculations could not accurately reflect the preoperative reliability of the Penetration-Aspiration Scale.

The results are clearer for the FEES assessment using the Swallowing Performance Scale. Kendall's W, which is useful with the type of ordinal data studied here, showed fair and extremely strong reliability preoperatively and postoperatively, respectively. The results for ICC were less impressive, but it has been suggested that ICC can be strongly affected by prevalence and variability, much like κ .²⁶ It is likely that a wider range of scores provided across the Swallowing Performance Scale would have improved the ICC results. Nonetheless, the Kendall's W data on the Swallowing Performance Scale are similar to previously reported results, such as those of Nienstedt et al.¹⁴

The present study did not address any of the alternative means of assessing swallowing function, such as subjective assessments like the MDADI. As with the FEES test using the Penetration-Aspiration Scale, the MDADI, including Swedish and Italian versions, has proven to be reliable and valid in numerous clinical evaluations.^{30,37,38} However, like the FEES test before the present study, it has not yet been shown to be reliable and valid in the assessment of patients undergoing ACDF surgery.³⁰

Some limitations of the present study call for further research. A study involving patients with greater variability along the Penetration-Aspiration Scale would be particularly beneficial. Such a study could also be expanded to include questionnaire assessments such as the MDADI.³⁰ In addition, although the present study specifically assessed the reliability of the FEES test in patients undergoing revision ACDF, it did not assess the validity of this diagnostic tool.

CONCLUSIONS

Our data show that the FEES test interpreted through the Penetration-Aspiration Scale and the Swallowing Performance Scale appears to be a reliable tool for assessing postoperative dysphagia in patients undergoing revisionary ACDF surgery and can serve as a useful test for exploring this important outcome in this patient population.

REFERENCES

- Leder SB. Fiberoptic endoscopic evaluation of swallowing in patients with acute traumatic brain injury. *J Head Trauma Rehabil.* 1999;14:448-453.
- Beutler WJ, Sweeney CA, Connolly PJ. Recurrent laryngeal nerve injury with anterior cervical spine surgery risk with laterality of surgical approach. *Spine (Phila Pa 1976).* 2001;26:1337-1342.
- Paniello RC, Martin-Bredahl KJ, Henkener LJ, Riew KD. Preoperative laryngeal nerve screening for revision anterior cervical spine procedures. *Ann Otol Rhinol Laryngol.* 2008;117:594-597.
- Mukherjee R, Müller M, Amstad H, et al. [Dysphonia and dysphagia after anterior cervical spine surgery]. *HNO.* 2014;62:575-581 [in German].
- Erwood MS, Hadley MN, Gordon AS, Carroll WR, Agee BS, Walters BC. Recurrent laryngeal nerve injury following reoperative anterior cervical discectomy and fusion: a meta-analysis. *J Neurosurg Spine.* 2016;25:198-204.
- Chotai S, Sivaganesan A, Parker SL, McGirt MJ, Devin CJ. Patient-specific factors associated with dissatisfaction after elective surgery for degenerative spine diseases. *Neurosurgery.* 2015;77:157-163 [discussion: 163].
- Cho SK, Lu Y, Lee DH. Dysphagia following anterior cervical spinal surgery: a systematic review. *Bone Joint J.* 2013;95-B:868-873.
- Gerek M, Atalay A, Cekin E, Ciyiltepe M, Ozkaptan Y. [The effectiveness of fiberoptic endoscopic swallow study and modified barium

- swallow study techniques in diagnosis of dysphagia]. *Kulak Burun Boğaz İhtis Derg.* 2005;15:103-111 [in Turkish].
9. Hiss SG, Postma GN. Fiberoptic endoscopic evaluation of swallowing. *Laryngoscope.* 2003;113:1386-1393.
 10. Robbins J, Coyle J, Rosenbek J, Roecker E, Wood J. Differentiation of normal and abnormal airway protection during swallowing using the Penetration-Aspiration Scale. *Dysphagia.* 1999;14:228-232.
 11. Salama JK, Stenson KM, List MA, et al. Characteristics associated with swallowing changes after concurrent chemotherapy and radiotherapy in patients with head and neck cancer. *Arch Otolaryngol Head Neck Surg.* 2008;134:1060-1065.
 12. Butler SG, Markley L, Sanders B, Stuart A. Reliability of the Penetration-Aspiration Scale with flexible endoscopic evaluation of swallowing. *Ann Otol Rhinol Laryngol.* 2015;124:480-483.
 13. Hey C, Pluschinski P, Pajunk R, et al. Penetration-Aspiration: is their detection in FEES reliable without video recording? *Dysphagia.* 2015;30:418-422.
 14. Nienstedt JC, Müller F, Nießen A, et al. Narrow band imaging enhances the detection rate of penetration and aspiration in FEES. *Dysphagia.* 2017;32:443-448.
 15. Rosenbek JC, Robbins JA, Roecker EB, Coyle J, Wood JL. A penetration-aspiration scale. *Dysphagia.* 1996;11:93-98.
 16. Karanicolos PJ, Bhandari M, Kreder H, et al. Evaluating agreement: conducting a reliability study. *J Bone Joint Surg Am.* 2009;91(suppl 3):99-106.
 17. Erwood MS, Walters BC, Connolly TM, et al. Voice and swallowing outcomes following reoperative anterior cervical discectomy and fusion with a 2-team surgical approach. *J Neurosurg Spine.* 2018;28:140-148.
 18. Walter SD, Eliasziw M, Donner A. Sample size and optimal designs for reliability studies. *Stat Med.* 1998;17:101-110.
 19. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics.* 1977;33:159-174.
 20. Byrt T, Bishop J, Carlin JB. Bias, prevalence and kappa. *J Clin Epidemiol.* 1993;46:423-429.
 21. Kendall MG, Gibbons JD. *Rank Correlation Methods.* 5th ed. London: Edward Arnold; 1990.
 22. Chen B, Zaebe D, Seel L. A macro to calculate kappa statistics for categorizations by multiple raters. Presented at the: Proceedings of the 30th SAS User Group International Conference. April 2005, Philadelphia, PA. Available at: <https://support.sas.com/resources/papers/proceedings/proceedings/sugi30/155-30.pdf>. Accessed August 7, 2017.
 23. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull.* 1979;86:420-428.
 24. McGraw KO, Wong SP. Forming inferences about some intraclass correlation coefficients. *Psychol Methods.* 1996;1:30-46.
 25. McHugh ML. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb).* 2012;22:276-282.
 26. Ten Cate DF, Luime JJ, Hazes JMW, Jacobs JWG, Landewé R. Does the intraclass correlation coefficient always reliably express reliability? Comment on the article by Cheung et al. *Arthritis Care Res (Hoboken).* 2010;62:1357-1358 [author reply: 1358].
 27. Thomas RL, Zidan MA, Slovis TL. What you need to know about statistics, part I: validity of diagnostic and screening tests. *Pediatr Radiol.* 2015;45:146-152.
 28. Zidan M, Thomas RL, Slovis TL. What you need to know about statistics, part II: reliability of diagnostic and screening tests. *Pediatr Radiol.* 2015;45:317-328.
 29. Stenson KM, MacCracken E, List M, et al. Swallowing function in patients with head and neck cancer prior to treatment. *Arch Otolaryngol Head Neck Surg.* 2000;126:371-377.
 30. Chen AY, Frankowski R, Bishop-Leone J, et al. The development and validation of a dysphagia-specific quality-of-life questionnaire for patients with head and neck cancer: the MD Anderson dysphagia inventory. *Arch Otolaryngol Head Neck Surg.* 2001;127:870-876.
 31. Fattori B, Siciliano G, Mancini V, et al. Dysphagia in amyotrophic lateral sclerosis: relationships between disease progression and fiberoptic endoscopic evaluation of swallowing. *Auris Nasus Larynx.* 2017;44:306-312.
 32. Leder SB. Serial fiberoptic endoscopic swallowing evaluations in the management of patients with dysphagia. *Arch Phys Med Rehabil.* 1998;79:1264-1269.
 33. Bader CA, Niemann G. [Dysphagia in children and young persons. The value of fiberoptic endoscopic evaluation of swallowing]. *HNO.* 2008;56:397-401 [in German].
 34. Leder SB, Karas DE. Fiberoptic endoscopic evaluation of swallowing in the pediatric population. *Laryngoscope.* 2000;110:1132-1136.
 35. Hey C, Pluschinski P, Zaretsky Y, et al. [Penetration-Aspiration Scale according to Rosenbek. Validation of the German version for endoscopic dysphagia diagnostics]. *HNO.* 2014;62:276-281 [in German].
 36. Feinstein AR, Cicchetti DV. High agreement but low kappa: I. The problems of two paradoxes. *J Clin Epidemiol.* 1990;43:543-549.
 37. Carlsson S, Rydén A, Rudberg I, Bove M, Bergquist H, Finizia C. Validation of the Swedish M. D. Anderson Dysphagia Inventory (MDADI) in patients with head and neck cancer and neurologic swallowing disturbances. *Dysphagia.* 2012;27:361-369.
 38. Schindler A, Borghi E, Tiddia C, Ginocchio D, Felisati G, Ottaviani F. Adaptation and validation of the Italian MD Anderson dysphagia inventory (MDADI). *Rev Laryngol Otol Rhinol (Bord).* 2008;129:97-100.

Conflict of interest statement: The authors declare that the article content was composed in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Received 7 March 2019; accepted 4 June 2019

Citation: World Neurosurg. (2019) 130:e199-e205. <https://doi.org/10.1016/j.wneu.2019.06.028>

Journal homepage: www.journals.elsevier.com/world-neurosurgery

Available online: www.sciencedirect.com

1878-8750/\$ - see front matter Published by Elsevier Inc.