CrossMark

ORIGINAL ARTICLE

The Yale Pharyngeal Residue Severity Rating Scale: An Anatomically Defined and Image-Based Tool

Paul D. Neubauer[1] · Alfred W. Rademaker[2] · Steven B. Leder[1]

Abstract The Yale Pharyngeal Residue Severity Rating Scale was developed, standardized, and validated to provide reliable, anatomically defined, and image-based assessment of post-swallow pharyngeal residue severity as observed during fiberoptic endoscopic evaluation of swallowing (FEES). It is a five-point ordinal rating scale based on residue location (vallecula and pyriform sinus) and amount (none, trace, mild, moderate, and severe). Two expert judges reviewed a total of 261 FEES evaluations and selected a no residue exemplar and three exemplars each of trace, mild, moderate, and severe vallecula and pyriform sinus residue. Hard-copy color images of the no residue, 12 vallecula, and 12 pyriform sinus exemplars were randomized by residue location for hierarchical categorization by 20 raters with a mean of 8.3 years of experience (range 2–27 years) performing and interpreting FEES. Severity ratings for all images were performed by the same 20 raters, 2 weeks apart, and with the order of image presentations randomized. Intra-rater test–retest reliability, inter-rater reliability, and construct validity were determined by pooled multi-category multi-rater kappa statistics. Residue ratings were excellent for intra-rater reliability for vallecula (kappa = 0.957 ± 0.014) and pyriform sinus (kappa = 0.854 ± 0.021); very good to excellent for inter-rater reliability for vallecula (kappa = 0.868 ± 0.011) and pyriform sinus (kappa = 0.751 ± 0.011); and excellent for validity for vallecula (kappa = 0.951 ± 0.014) and pyriform sinus (kappa = 0.908 ± 0.017). Clinical uses include accurate classification of vallecula and pyriform sinus residue severity patterns as none, trace, mild, moderate, or severe for diagnostic purposes, determination of functional therapeutic change, and precise dissemination of shared information. Scientific uses include tracking outcome measures, demonstrating efficacy of interventions to reduce pharyngeal residue, investigating morbidity and mortality in relation to pharyngeal residue severity, and improving training and accuracy of FEES interpretation by students and clinicians. The Yale Pharyngeal Residue Severity Rating Scale is a reliable, validated, anatomically defined, and image-based tool to determine residue location and severity based on FEES.

Introduction

Pharyngeal residue, defined as pre-swallow secretions and post-swallow food residue in the pharynx not entirely cleared by a swallow, is a clinical predictor of prandial aspiration [1]. An accurate description of pharyngeal residue severity is an important but difficult clinical challenge

✉ Steven B. Leder
steven.leder@yale.edu

Paul D. Neubauer
paul.neubauer@yale.edu

Alfred W. Rademaker
rademaker@northwestern.edu

1 Section of Otolaryngology, Department of Surgery, Yale School of Medicine, P.O. Box 208041, New Haven, CT 06520-8041, USA

2 Department of Preventive Medicine, Biostatistics Core Facility, The Robert H. Lurie Comprehensive Cancer Center, Northwestern University Feinberg School of Medicine, Chicago, IL, USA

 Springer

[2]. Pharyngeal residue occurs in either the valleculae (spaces between the base of tongue and epiglottis) or the pyriform sinuses (spaces formed on both sides of the pharynx between the fibers of the inferior pharyngeal constrictor muscle and the sides of the thyroid cartilage and lined by orthogonally directed fibers of the palatopharyngeus muscle and pharyngobasilar fascia) [3].

Different types of scales have attempted to classify pharyngeal residue but none have demonstrated the combination of adequate reliability, interpretive validity, and ease of administration to be clinically useful. Scale examples are as follows: 1. Binary (presence/absence) [4]; 2. Ordinal (to capture progressively increasing amounts) [1, 5–8]; 3. Estimation (amount of observed residue as an estimate of the percentage of the original bolus) [9–12]; and 4. Quantification (computer-based image analysis) [2, 13]. The sole purpose of all of these scales is to rate pharyngeal residue severity. These scales do not determine why residue occurs or ascertain the timing of residue occurrence during swallowing. No scale, to date, has provided in vivo, anatomically correct, image-based exemplars of graduated pharyngeal residue severity ratings against which clinicians can match their clinical judgments.

Fiberoptic endoscopic evaluation of swallowing (FEES) [14, 15] is a recognized, validated, and widely used technique to assess the pharyngeal phase of swallowing in order to diagnose dysphagia, recommend oral diets, and implement appropriate rehabilitation interventions; all with the goal of promoting safe and efficient swallowing [12, 16–19]. The endoscopist is alert to pre-swallow pooled secretions and post-swallow food residue in the pharynx. FEES has been shown to be more sensitive in identifying pharyngeal residue when compared to the videofluoroscopic swallow study (VFSS) [12, 19]. Since pharyngeal residue is an important predictor of swallowing success [1], it is important to ascertain residue severity in the valleculae and pyriform sinuses. However, to date, pharyngeal residue severity has not been described using an objective, anatomically correct, image-based, reliable, and validated rating scale based on FEES.

Standardized evaluation of depth of laryngeal penetration and aspiration has only been reported with the penetration-aspiration scale (PAS) [20]. The PAS is an 8-point scale ranging from "1—material does not enter the airway" to "8—material enters the airway, passes below the vocal folds, and no effort is made to eject." The PAS was validated using VFSS and does not rate pharyngeal residue.

The presence of pre-swallow pooled secretions and post-swallow food residue in the *laryngeal vestibule* is an important sign of potential poor swallowing performance and increased aspiration risk later during FEES. Pooled secretions in the laryngeal vestibule were highly predictive of prandial aspiration in adults [1] and correlated with aspiration pneumonia in children [21]. However, determining bolus volume patterns in the laryngeal vestibule poses a particular problem as only trace and mild occur before caudal bolus flow results in aspiration. Therefore, the focus of the present study is solely on *pharyngeal residue*.

There is no objective, anatomically defined, image-based, reliable, and validated tool to rate severity of residue in the valleculae and pyriform sinuses during FEES. It would be advantageous for clinicians to be able to reliably determine, monitor, and share their patients' pharyngeal residue patterns. The purpose of this study was to develop, standardize, and validate the Yale Pharyngeal Residue Severity Rating Scale with the goal of providing objective, anatomically defined, image-based, reliable, and validated pharyngeal residue severity ratings based upon FEES.

Methods

Subjects

This study was approved by the Human Investigation Committee, Yale School of Medicine. Non-identified adult FEES evaluations performed at Yale-New Haven Hospital during 2013–2014 were used. Gender, age, ethnicity, and diagnosis were deemed not to influence the review of images by the raters.

Fiberoptic Endoscopic Evaluation of Swallowing (FEES)

The standard FEES protocol was followed with slight modifications [14, 15]. Briefly, each naris was examined visually and the scope passed through the most patent naris without administration of a topical anesthetic or vasoconstrictor to the nasal mucosa, thereby eliminating any potential adverse anesthetic reaction and assuring the endoscopist of a safe physiologic examination [22]. The base of tongue, pharynx, and larynx was viewed and swallowing was evaluated directly with six food boluses of approximately 5–10 cc volume each. Patients were encouraged to feed themselves, with assistance as needed, i.e., liquid with a straw or cup and puree with a spoon. All patients were allowed to swallow spontaneously, i.e., without a verbal command to swallow [23]. FEES equipment consisted of a distal chip flexible fiberoptic rhinolaryngoscope (KayPentax, Lincoln Park, NJ 07035, model VNL-117OK), light source (KayPentax, model EPK-1000), and a digital swallow workstation (KayPentax, model 7200).

The first food challenge consisted of three boluses of puree consistency (yellow pudding) followed by three thin

liquid boluses (white, fat free, skim milk), as these colors have excellent contrast with pharyngeal and laryngeal mucosa [24]. A solid food challenge, i.e., graham cracker, was given only if the patient was dentate.

Severity Rating Definitions

Definitions were anatomically defined, image-based, and used a five-point ordinal rating scale that encompassed the full range of severity ratings, i.e., none, trace, mild, moderate, and severe, for both the vallecula and pyriform sinus locations (Appendix).

Image Selection Process

In the absence of a criterion standard, two expert judges were considered the best referent standards. These two judges, with a combined 26 years of performing and interpreting FEES, reviewed a total of 261 FEES evaluations. All images were stored on a digital swallow workstation allowing for frame-by-frame editing. No audio cues were used. A total of 101 potential images were selected based on adequate image quality and severity criteria as defined in the Appendix. Consensus agreement allowed for selection of 25 potential final images, i.e., a no residue exemplar and three exemplars each of trace, mild, moderate, and severe vallecula and pyriform sinus residue. Hard-copy color images of the no residue, 12 vallecula, and 12 pyriform sinus images were randomized by residue location for hierarchical categorization by 20 raters.

Raters

A total of 20 raters trained at 18 different institutions from around the world participated, i.e., otolaryngology residents ($n = 11$), attending otolaryngologists ($n = 5$), speech-language pathologists ($n = 3$), and physician assistant ($n = 1$). The raters had different durations of experience in performing and interpreting FEES evaluations (mean 8.3 years, range 2–27 years).

Raters were grouped by years of FEES experience and training status. Years of experience indicated that ten raters had ≤ 4 years (mean 2.8 years, range 2–4 years) and ten raters had ≥ 5 years (mean 13.4 years, range 5–27 years). Training was done once, with random assignment of ten raters to receive and ten raters not to receive pre-rating training in determining vallecula and pyriform sinus pharyngeal residue severity ratings. Training included written definitions, visual depictions, verbal explanations, and clarifying questions/answers of the severity ratings. No training was limited to only written definitions and visual depictions of the severity ratings.

Reliability Testing

Intra-rater test–retest reliability, inter-rater reliability, and construct validity for severity ratings for all images were performed by the same two expert judges and 20 raters, 2 weeks apart, and with the order of image presentations randomized. This allowed for selection of the best representative exemplar in each severity rating, i.e., none, trace, mild, moderate, and severe.

Statistics

Analyses were done separately for vallecula and pyriform sinus locations. Therefore, there was a total of 260 ratings (20 raters rated 13 images) for each location at each time point. Kappa statistics and their standard errors were used to assess the extent of intra- and inter-rater reliability and construct validity [25]. *Intra-rater reliability* was calculated by pooling the 260 paired ratings and calculating a weighted kappa [25, p. 223], weighted by the degree of disagreement, with comparison of the same image 2 weeks apart. A similar analysis was done to assess *construct validity* by comparing the initial ratings with the criterion standard ratings from the two expert judges. *Inter-rater reliability* was calculated using a multi-rater kappa [25, p. 226] where the extent of agreement across raters was calculated for each of the five categories (none, trace, mild, moderate, and severe) followed by calculation of a weighted average of these category specific agreements, weighted by the number of ratings for each category. Weights were 1/13 for no residue and 3/13 for each of trace, mild, moderate, and severe residue. Kappa statistics ± standard error (se) is reported. Kappa statistics were compared across subsets of years of experience and training using Z-statistics.

Results

Intra- and inter-rater reliability was 100 % for the two expert judges based on the rating of the 25 potential final scale images, i.e., a no residue exemplar and three examples each of trace, mild, moderate, and severe vallecula and pyriform sinus residue.

The Yale Pharyngeal Residue Severity Rating Scale demonstrated excellent overall intra-rater kappa statistics for both locations, specifically, 1. Intra-rater reliability for vallecula (0.957 ± 0.014) and pyriform sinus (0.854 ± 0.021); 2. Inter-rater reliability for vallecula (0.868 ± 0.011) and pyriform sinus (0.751 ± 0.011); and 3. Construct validity for vallecula (0.951 ± 0.014) and pyriform sinus (0.908 ± 0.017) (Table 1).

Table 1 Intra-rater test-retest reliability, inter-rater reliability, and construct validity kappa statistics (standard error) for vallecula and pyriform sinus residue ratings across all raters ($n = 20$)

Location*	Kappa (se)
Intra-rater reliability	
Vallecula	0.957 (±0.014)
Pyriform sinus	0.854 (±0.021)
Inter-rater reliability	
Vallecula	0.868 (±0.011)
Pyriform sinus	0.751 (±0.011)
Construct validity	
Vallecula	0.951 (±0.014)
Pyriform sinus	0.908 (±0.017)

* Total of 260 ratings for each location

Table 2 Intra-rater test-retest reliability and inter-rater reliability kappa statistics (standard error) for vallecula and pyriform sinus residue ratings based on years of experience ≤4 years ($n = 10$) versus ≥5 years ($n = 10$)

Location*	Kappa (se)	Location	Kappa (se)
Intra-rater reliability			
Vallecula		Pyriform sinus	
≤4 years	0.968 (±0.017)	≤4 years	0.823 (±0.032)
≥5 years	0.946 (±0.021)	≥5 years	0.881 (±0.028)
Z = 0.82		Z = 1.36	
p value 0.38		*p* value 0.17	
Inter-rater reliability			
Vallecula		Pyriform sinus	
≤4 years	0.921 (±0.022)	≤4 years	0.648 (±0.022)
≥5 years	0.816 (±0.022)	≥5 years	0.886 (±0.022)
Z = 3.38		Z = 7.68	
p < 0.001		*p* < 0.001	

* Total of 130 ratings for each subset

Table 3 Intra-rater test-retest reliability and inter-rater reliability kappa statistics (standard error) for vallecula and pyriform sinus residue based on training ($n = 10$) versus no training ($n = 10$)

Location*	Kappa (se)	Location	Kappa (se)
Intra-rater reliability			
Vallecula		Pyriform sinus	
Training	0.989 (±0.008)	Training	0.838 (±0.033)
No training	0.924 (±0.026)	No training	0.870 (±0.027)
Z = 2.39		Z = 0.75	
p value 0.02		*p* value 0.45	
Inter-rater reliability			
Vallecula		Pyriform sinus	
Training	0.961 (±0.022)	Training	0.805 (±0.022)
No training	0.777 (±0.022)	No training	0.680 (±0.022)
Z = 5.93		Z = 4.03	
p < 0.001		*p* < 0.001	

* Total of 130 ratings for each subset

Table 4 Construct validity kappa statistics (standard error) for vallecula and pyriform sinus residue ratings based on years of experience and training

Location	Kappa (se)	Location	Kappa (se)
Years of experience			
Vallecula		Pyriform sinus	
≤4 years	0.968 (±0.017)	≤4 years	0.848 (±0.031)
≥5 years	0.935 (±0.023)	≥5 years	0.967 (±0.013)
Z = 1.15		Z = 3.54	
p value 0.25		*p* value <0.001	
Training			
Vallecula		Pyriform sinus	
Training	0.989 (±0.008)	Training	1.000 (0.000)
No training	0.913 (±0.027)	No training	0.881 (±0.028)
Z = 2.70		Z = 4.25	
p value 0.007		*p* value <0.001	

Intra-rater kappa statistics were between 0.823 ± 0.032 and 0.969 ± 0.017 dependent upon the location of residue (vallecula or pyriform sinus) and raters' years of experience (Table 2). No differences by years of experience for intra-rater reliability for vallecula ($p = 0.38$) and pyriform sinus ($p = 0.17$) kappas were found. Inter-rater reliability for years of experience was not consistent, i.e., ≤4 years had higher kappas for vallecula ($p < 0.001$) but lower kappas for pyriform sinus ($p < 0.001$).

Intra-rater kappa statistics were between 0.838 ± 0.033 and 0.989 ± 0.008 dependent upon the location of residue (vallecula or pyriform sinus) and training versus no training (Table 3). A difference was found in favor of training with higher vallecula kappas ($p = 0.02$) but no difference was found for pyriform sinus kappas ($p = 0.45$). Inter-rater kappa statistics were between 0.680 ± 0.022 and 0.961 ± 0.022

dependent upon the location of residue (vallecula or pyriform sinus) and training resulted in higher kappas for both locations ($p < 0.001$).

Construct validity kappa statistics were between 0.848 ± 0.031 and 1.000 dependent upon the location of residue (vallecula or pyriform sinus) and either years of experience or training status (Table 4). More years of experience had higher kappa values for pyriform sinus ($p = 0.001$) and there was no difference by years of experience for vallecula ($p = 0.25$). Training again resulted in higher kappas for both vallecula ($p = 0.007$) and pyriform sinus ($p = 0.001$).

Table 5 Inter-rater reliability and construct validity kappa statistics (standard error) for re-randomized vallecula and pyriform sinus residue images rated two weeks later based on years of experience ≤4 years (n = 10) versus ≥5 years (n = 10) and training (n = 10) versus no training (n = 10)

Location	Kappa (se)	Location	Kappa (se)
Inter-rater reliability			
Years of experience			
Vallecula		Pyriform sinus	
≤4 years	1.000 (±0.022)	≤4 years	0.670 (±0.022)
≥5 years	0.960 (±0.022)	≥5 years	0.774 (±0.022)
Z = 1.29		Z = 3.35	
p value 0.23		*p* value <0.001	
Training			
Vallecula		Pyriform sinus	
Training	1.000 (±0.022)	Training	0.698 (±0.022)
No training	0.961 (±0.022)	No training	0.726 (±0.022)
Z = 1.26		Z = 0.90	
p value 0.21		*p* value 0.32	
Construct validity			
Years of experience			
Vallecula		Pyriform sinus	
≤4 years	1.000 (0.000)	≤4 years	0.870 (±0.027)
≥5 years	0.989 (±0.008)	≥5 years	0.913 (±0.024)
Z = 1.38		Z = 1.19	
p value 0.20		*p* value 0.23	
Training			
Vallecula		Pyriform sinus	
Training	1.000 (0.000)	Training	0.881 (±0.028)
No training	0.989 (±0.008)	No training	0.903 (±0.023)
Z = 1.38		Z = 0.61	
p value 0.17		*p* value 0.55	

Inter-rater reliability kappa statistics for re-randomized images rated 2 weeks later were between 0.670 ± 0.022 and 1.000 ± 0.022 for years of experience and between 0.698 ± 0.022 and 1.000 ± 0.002 for training (Table 5). More years of experience had higher kappa values for pyriform sinus ($p < 0.001$) and there was no difference by years of experience for vallecula ($p = 0.23$). Training did not result in higher kappa values for both vallecula ($p = 0.21$) and pyriform sinus ($p = 0.32$). Construct validity kappa statistics were between 0.870 ± 0.027 and 1.000 dependent upon the location of residue (vallecula or pyriform sinus) and either years of experience or training status. More years of experience did not result in higher kappa values for either vallecula ($p = 0.20$) or pyriform sinus ($p = 0.23$). Training did not result in higher kappa values for either vallecula ($p = 0.17$) or pyriform sinus ($p = 0.55$).

The single image with the greatest inter-rater agreement for each residue severity level, i.e., none, trace, mild, moderate, and severe, and for each location, i.e., vallecula (Fig. 1) and pyriform sinus (Fig. 2) became the chosen exemplar for inclusion in the Yale Pharyngeal Residue Severity Rating Scale.

Discussion

The Yale Pharyngeal Residue Severity Rating Scale has achieved its stated goal of providing reliable and valid information regarding the location and severity of pharyngeal residue observed during FEES. Vallecula and pyriform sinus residue severity ratings showed overall excellent intra-rater reliability, inter-rater agreement, and construct validity. Importantly, repeat ratings 2 weeks later of the same but re-randomized images found that neither years of experience nor training status resulted in higher validity kappa values for vallecula and pyriform sinus ratings. Therefore, proficiency in the use of the Yale Pharyngeal Residue Severity Rating Scale is readily achievable in a short period of time by clinicians from different specialty areas and with different levels of expertise.
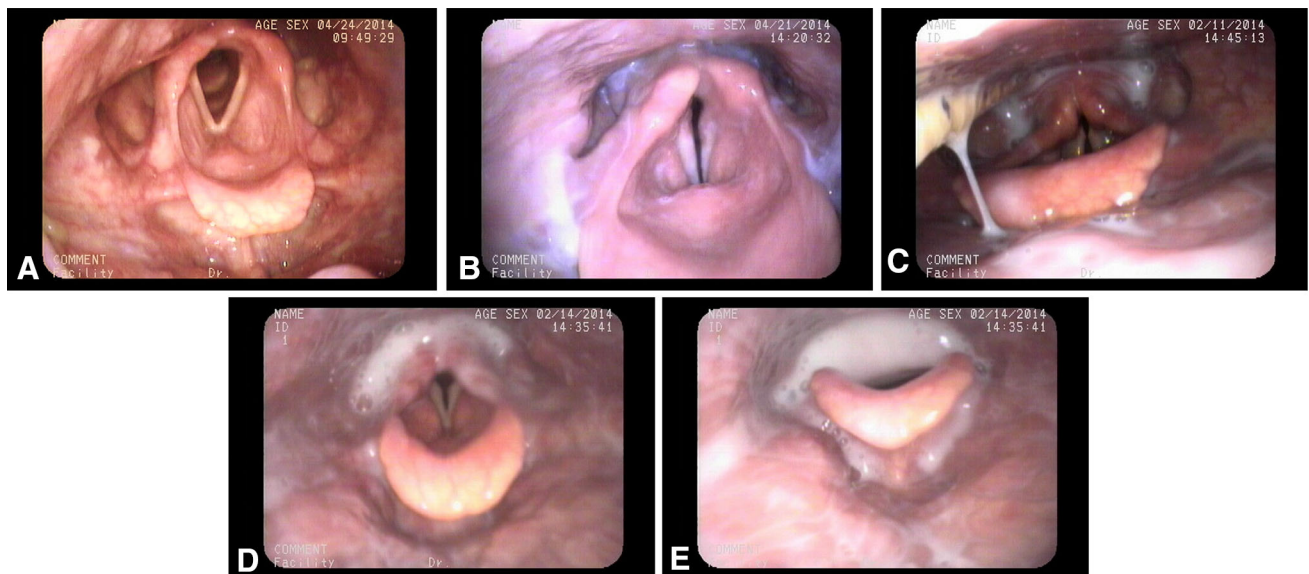
The sole purpose of the Yale Pharyngeal Residue Severity Rating Scale is to allow clinicians and researchers rate post-swallow vallecular and pyriform sinus residue severity. Consistent with all other pharyngeal residue rating scales [1, 2, 4–13], the Yale Pharyngeal Residue Severity Rating Scale does not determine why residue occurs or ascertain the timing of residue occurrence during swallowing. Since all patients have unique swallowing characteristics, it is up to the clinician to determine the why and when of residue occurrence during swallowing. The superiority of the Yale Pharyngeal Residue Severity Rating Scale is due to its anatomically defined and image-based construction resulting in excellent validity, easy administration, and accurate interpretation by clinicians with a wide range of FEES experience, and generalizability to all individuals.

The utility, versatility, and efficacy of the Yale Pharyngeal Residue Severity Scale are easily demonstrated. For example, a representative pre-therapy swallow receives a severe vallecula residue severity rating (anatomically defined as the vallecula filled up to the epiglottic rim and with a corresponding image). An intervention strategy, such as effortful swallow or double-swallow, is implemented for a set period of time and a representative post-therapy swallow receives a mild vallecular residue severity rating (anatomically defined as mild pooling with epiglottic ligament visible and with a corresponding image). The clinician can now document efficacy of a specific treatment intervention and either stop, continue, or change strategies.
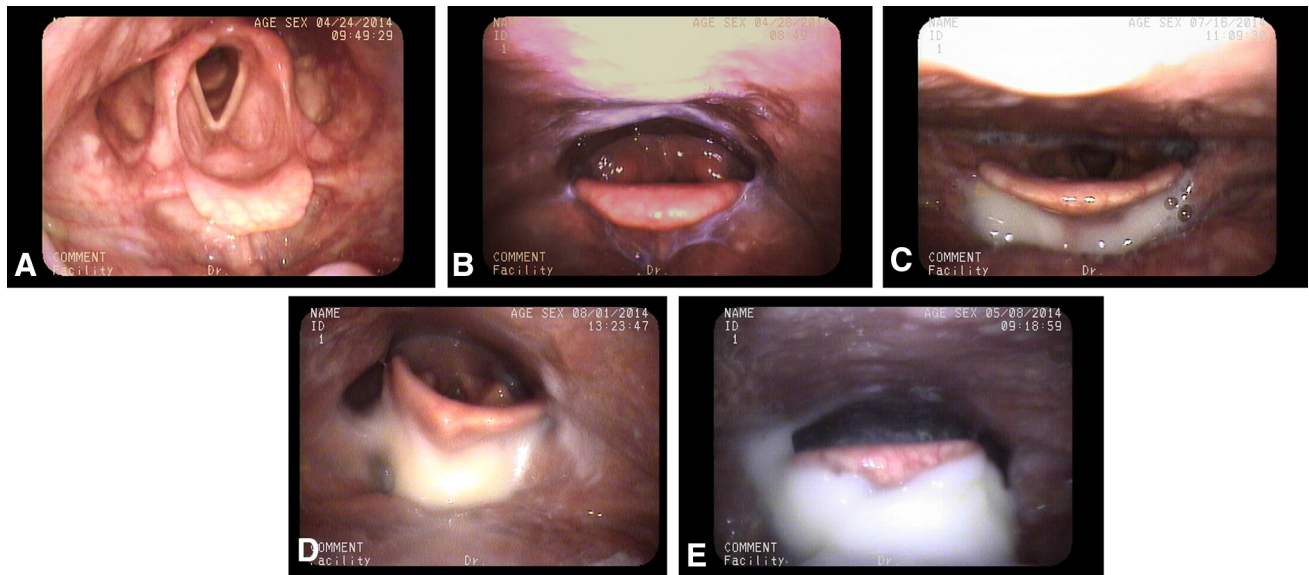
Fig. 1 The vallecula images with the greatest inter-rater agreement for each residue level: **a** none; **b** trace; **c** mild; **d** moderate; and **e** severe

Fig. 2 The pyriform sinus images with the greatest inter-rater agreement for each residue level: **a** none; **b** trace; **c** mild; **d** moderate; and **e** severe

Prior to the development and validation of the Yale Pharyngeal Residue Severity Rating Scale, objective documentation of therapeutic interventions was not possible.

The Yale Pharyngeal Residue Rating Scale works well for any swallow, whether it is the first, subsequent clearing, or last swallow. The clinician simply has to match their chosen swallow with its scale mate. In this way, it is possible to determine if spontaneous or volitional clearing swallows or a throat clearing maneuver is actually helpful in reducing the amount of residue in the vallecula and pyriform sinuses. Since an important therapeutic goal is to

aid pharyngeal clearing [1], this information can guide intervention strategies and promote safer swallowing. For example, it is now possible to determine objectively if drinking a small liquid bolus after a puree/solid bolus, an effortful swallow, a double-swallow/bolus, a head turn to left or right, and a chin tuck are successful in reducing residue in the vallecula and pyriform sinus.

Since the anatomical definitions used by the Yale Pharyngeal Residue Scale are discrete, i.e., not continuous, and image-based, the severity rating is not affected by age, gender, or body habitus. For example, mild vallecula

residue is defined as "epiglottic ligament visible." The shape and size of the vallecula are unimportant. As long as the epiglottic ligament is visible, the severity rating is mild residue. This generalizability makes it possible to determine pharyngeal residue severity for any given individual.

The Yale Pharyngeal Residue Severity Rating Scale can be used for both clinical advantages and research opportunities. Clinically, clinicians can now accurately classify vallecula and pyriform sinus residue severity as none, trace, mild, moderate, or severe for diagnostic purposes, determination of functional therapeutic change, and precise dissemination of shared information. Future research uses include tracking outcome measures for clinical trials investigating various swallowing interventions, demonstrating efficacy of specific interventions to reduce pharyngeal residue, determining morbidity and mortality associated with pharyngeal residue severity in different patient populations, and improving the training and accuracy of FEES interpretation by students and clinicians.

Conclusions

The Yale Pharyngeal Residue Severity Rating Scale is a reliable, validated, anatomically defined, and image-based tool to determine residue location and severity based on FEES. Proficiency can be readily achieved with minimal training and at high levels of intra- and inter-rater reliability and construct validity. Clinical uses include, but are not limited to, accurate classification of vallecula and pyriform sinus residue severity patterns as none, trace, mild, moderate, or severe for diagnostic purposes, determination of functional therapeutic change, and precise dissemination of shared information. Scientific uses include, but are not limited to, tracking outcome measures, demonstrating efficacy of interventions to reduce pharyngeal residue, investigating morbidity and mortality in relation to pharyngeal residue severity, and improving training and accuracy of FEES interpretation by students and clinicians. The Yale Pharyngeal Residue Severity Rating Scale is an important addition to the deglutologist's tool box and can be used with confidence for both clinical and research purposes.

Appendix

See Tables 6 and 7.

Table 6 Definitions for severity of vallecula residue

I	None	0 %	No residue
II	Trace	1–5 %	Trace coating of the mucosa
III	Mild	5–25 %	Epiglottic ligament visible
IV	Moderate	25–50 %	Epiglottic ligament covered
V	Severe	>50 %	Filled to epiglottic rim

Table 7 Definitions for severity of pyriform sinus residue

I	None	0 %	No residue
II	Trace	1–5 %	Trace coating of mucosa
III	Mild	5–25 %	Up wall to quarter full
IV	Moderate	25–50 %	Up wall to half full
V	Severe	>50 %	Filled to aryepiglottic fold

References

1. Murray J, Langmore SE, Ginsberg S, Dostie A. The significance of oropharyngeal secretions and swallowing frequency in predicting aspiration. Dysphagia. 1996;11:99–103.
2. Pearson WG, Molfenter SM, Smith ZM, Steele CM. Image-based measurement of post- swallow residue: the normalized residue ratio scale. Dysphagia. 2013;28:167–77.
3. Logemann J. Evaluation and treatment of swallowing disorders. 2nd ed. Austin: Pro-Ed; 1998.
4. Dejaeger E, Pelemans W, Ponette E, Joosten E. Mechanisms involved in postdeglutition retention in the elderly. Dysphagia. 1997;12:63–7.
5. Farneti D. Pooling score: an endoscopic model for evaluating severity of dysphagia. Acta Otorhinological Italica. 2008;28:135–40.
6. Tohara H, Nakane A, Murata S, Mikushi S, Ouchi Y, Wakasugi Y, Takashima M, Chiba Y, Uematsu H. Inter- and inter-rater reliability in fibroptic endoscopic evaluation of swallowing. J Oral Rehabil. 2010;37:884–91.
7. Kaneoka AS, Langmore SE, Krisciunas GP, Field K, Scheel R, McNally E, Walsh MJ, O'Dea MB, Cabral H. The Boston residue and clearance scale: preliminary reliability and validity testing. Folia Phoniatr Logop. 2014;65:312–7.
8. Donzelli J, Brady S, Wesling M, Craney M. Predictive value of accumulated oropharyngeal secretions for aspiration during video nasal endoscopic evaluation of the swallow. Ann Otol Rhinol. 2003;112:469–75.
9. Han TR, Paik NJ, Park JW. Quantifying swallowing function after stroke: a functional dysphagia scale based on videofluoroscopic studies. Arch Phys Med Rehabil. 2001;82:677–82.
10. Eisenhuber E, Schima W, Schober E, Pokieser P, Stadler A, Scharitzer M, Oschatz E. Videofluorosocpic assessment of patients with dysphagia: pharyngeal retention is a predictive factor for aspiration. AJR Am J Roentgenol. 2002;178:393–8.
11. Logemann JA, Williams RB, Rademaker A, Pauloski BR, Lazarus CL, Cook I. The relationship between observations and measures of oral and pharyngeal residue from videofluorography and scintigraphy. Dysphagia. 2005;20:226–31.
12. Kelly AM, Leslie P, Beale T, Payten C, Drinnan MJ. Fibreoptic endoscopic evaluation of swallowing and videofluoroscopy: does examination type influence perception of pharyngeal severity? Clin Otolaryngol. 2006;31:425–32.

13. Dyer JC, Leslie P, Drinnan MJ. Objective computer-based assessment of valleculae residue: is it useful? Dysphagia. 2008;23:7–15.
14. Langmore SE, Schatz K, Olsen N. Fiberoptic endoscopic examination of swallowing safety: a new procedure. Dysphagia. 1988;2:216–9.
15. Leder SB, Murray JT. Fiberoptic endoscopic evaluation of swallowing. Phys Med Rehabil Clin No Am. 2008;19:787–801.
16. Wu CH, Hsiao TY, Chen JC, Yeun-Chung C, Shiann-Yann L. Evaluation of swallowing safety with fiberoptic endoscope: comparison with videofluoroscopic technique. Laryngoscope. 1997;107:396–401.
17. Leder SB, Sasaki CT, Burrell MI. Fiberoptic endoscopic evaluation of dysphagia to identify silent aspiration. Dysphagia. 1998;13:19–21.
18. Leder SB, Karas DE. Fiberoptic endoscopic evaluation of swallowing in the pediatric population. Laryngoscope. 2000;110:1132–6.
19. Kelly AM, Drinnan MJ, Leslie P. Assessing penetration and aspiration: how do videofluoroscopy and fiberoptic endoscopic evaluation of swallowing compare? Laryngoscope. 2007;117:1723–7.
20. Rosenbek JC, Robbins JA, Roecker EB, Coyle JC, Wood JL. A penetration-aspiration scale. Dysphagia. 1996;11:93–8.
21. Link DT, Willging JP, Miller CK, Cotton RT, Rudolph CD. Pediatric laryngoscopic sensory testing during flexible endoscopic evaluation of swallowing: feasible and correlative. Ann Otol Rhinol Laryngol. 2000;109:899–905.
22. Leder SB, Ross DA, Briskin KB, Sasaki CT. A prospective, double-blind, randomized study on the use of topical anesthetic, vasoconstrictor, and placebo during transnasal flexible fiberoptic endoscopy. J Speech Lang Hear Res. 1997;40:1352–7.
23. Daniels SK, Schroeder MF, DeGeorge PC, Corey D, Rosenbek JC. Effects of verbal cue on bolus flow during swallowing. J Am Speech Lang Pathol. 2007;16:140–7.
24. Leder SB, Acton LA, Lisitano HL, Murray JT. Fiberoptic endoscopic evaluation of swallowing (FEES) with and without blue dyed food. Dysphagia. 2005;20:157–62.
25. Fleiss JL. Statistical Methods for Rates and Proportions. New York: Wiley; 1981.

Paul D. Neubauer   MD

Alfred W. Rademaker   PhD

Steven B. Leder   PhD, CCC-SLP