

Assessing Penetration and Aspiration: How Do Videofluoroscopy and Fiberoptic Endoscopic Evaluation of Swallowing Compare?

Annette M. Kelly, MSc; Michael J. Drinnan, PhD; Paula Leslie, PhD

Objectives/Hypothesis: We aimed to investigate whether the type of dysphagia examination (fiberoptic endoscopic evaluation of swallowing [FEES] or videofluoroscopy) influences the scoring of penetration and aspiration.

Study Design: Prospective, single-blind study.

Methods: Fifteen dysphagic participants were recruited and underwent one FEES and one videofluoroscopy examination, performed and recorded simultaneously. Fifteen independent raters from 12 centers scored penetration and aspiration from recordings using the Penetration Aspiration Scale. Raters were blind to participant details, the pairing of the FEES and videofluoroscopy recordings, and the other raters' scores. Interrater and intrarater reliability were analyzed using weighted kappa.

Results: The Penetration Aspiration Scale scores were significantly higher for the FEES recordings than for the videofluoroscopy recordings (ANOVA $P < .001$). The mean difference between the FEES and videofluoroscopy penetration aspiration scores for the same swallows was 1.15 points. Interrater and intrarater reliability ranged from 0.64 to 0.79 (weighted kappa).

Conclusions: Penetration aspiration is perceived to be greater (more severe) from FEES than videofluoroscopy images. The clinical implications are discussed.

Key Words: Deglutition, swallowing, dysphagia, fiberoptic endoscopic evaluation of swallowing, videofluoroscopy, penetration, aspiration.

Laryngoscope, 117:1723–1727, 2007

INTRODUCTION

Fiberoptic endoscopic evaluation of swallowing (FEES) and videofluoroscopy are complementary examinations used routinely to diagnose oropharyngeal swallowing disorders and guide dysphagia management. Both examinations are well established as clinical tools, but their respective strengths, limitations and health outcomes continue to be debated in the literature.^{1–4}

Clinical judgment of penetration or aspiration and the patient's response to it are the most important factors in estimating the severity of oropharyngeal dysphagia and the risk of related complications such as aspiration pneumonia. Clinicians base their recommendations for modified oral intake or non-oral feeding on the results of FEES and videofluoroscopy examinations, and the evidence linking clinical diagnosis of aspiration to increased pneumonia risk supports this clinical practice.^{5,6} The perception of aspiration risk from clinical dysphagia assessments is a key factor underpinning oral intake recommendations. These decisions have potential implications for health outcomes and increased care costs related to tube feeding and delayed discharge.

Several studies have compared FEES and videofluoroscopy for detecting penetration and aspiration,^{7–11} but only one conducted the examinations simultaneously to enable comparison of the same swallows.¹⁰ All raters scored penetration and aspiration as present or absent rather than differentiating between degrees of entry of food or fluid into the airway. Despite the subjective nature of FEES and videofluoroscopy, only one of the five studies provided interrater reliability information¹⁰ and these data were limited. Therefore it remains unknown whether clinicians are influenced by the use of FEES or videofluoroscopy in diagnosing clinically

From the Head and Neck Tumour Centre (A.M.K.), University College Hospital, London, U.K., and the Department of Regional Medical Physics (M.J.D.), Freeman Hospital, Newcastle upon Tyne, U.K., and the Department of Communication Science and Disorders (P.L.), University of Pittsburgh, Pittsburgh, Pennsylvania, U.S.A.

Editor's Note: This Manuscript was accepted for publication May 22, 2007.

Supported by the Rhinology and Laryngology Research Fund, Institute of Laryngology, University College London, U.K.

Presented at the British Academic Conference of Otolaryngology, Birmingham, United Kingdom, July 5, 2006.

Send correspondence to Annette M. Kelly, Head and Neck Tumour Centre, University College Hospital, 1st Floor East, 250 Euston Road, London, NW1 2PG, U.K. E-mail: Annette.Kelly@uclh.nhs.uk

DOI: 10.1097/MLG.0b013e318123ee6a

relevant penetration and aspiration. In a previous study, we conducted simultaneous FEES and videofluoroscopy examinations and found that pharyngeal residue was perceived to be significantly greater when viewed from FEES recordings than from videofluoroscopy recordings.¹² The aim of this study was to determine whether the type of examination similarly influences the perception of penetration and aspiration.

MATERIALS AND METHODS

The methods used in this study have been described in detail previously.¹² The analysis was replicated precisely in this study with the exception of the outcome measure. The methods are briefly described here.

Ethics

The Royal Free Hampstead NHS Trust Local Ethics Committee approved this study. Participants were given an information sheet at least 24 hours prior to giving informed, written consent.

Participants

Participants were recruited from referrals to speech and language therapy for the assessment of dysphagia. Those nil by mouth or judged to be at high risk of aspiration of all oral intake were excluded. The 15 participants were the same study group used in our previous study comparing FEES and videofluoroscopy. Their details are outlined in Table I.

FEES and Videofluoroscopy Examinations

All participants underwent one FEES and one videofluoroscopy examination, which were performed and recorded simulta-

neously. Two dysphagia specialist speech-language pathologists, a consultant radiologist and a radiographer performed the examinations. The participants swallowed two test boluses; 5 mL liquid (2.5 mL barium (Baritop 100, Sakai Chemical Industry Ltd, Osaka, Japan) and 2.5 mL water), from a cup and 15 mL yogurt (5 mL barium (Baritop 100) and 10 mL smooth yogurt), from a standard dessert spoon. Both test boluses were dyed with 1 mL of blue food dye to enable clear visualization of the bolus. Participants also swallowed an undyed 5 mL water bolus immediately after swallowing the liquid test bolus. The water bolus was given to rinse the pharynx and was not recorded. The yogurt bolus was swallowed immediately after the water bolus. All participants swallowed the three boluses in the same sequence (liquid test bolus, water, and yogurt test bolus).

Screening Fields and Recording Instrumentation

The endoscopic images of the oropharynx, hypopharynx and larynx were recorded on the Digital Swallowing Workstation (Kay Pentax Ltd, Montvale, NJ). The lateral videofluoroscopy screening field was bordered by lips, hard palate, cervical spine and cervical esophagus and the images were recorded on Super-VHS video (SVO-9620, Sony, Weybridge, U.K.). The endoscopic and fluoroscopic recordings were made simultaneously.

Recorded Video Clips

The 30 recordings (15 FEES and 15 videofluoroscopy) were converted into MPEG video segments and recorded onto two computer disks. The clips were labeled 1 to 30 and recorded in random sequence onto CD#1. The clips were then re-randomized and labeled 1 to 30 onto CD#2. Each video clip consisted of three segments: 1) the liquid swallow, 2) the word "yogurt" to alert the rater to the next bolus, 3) the yogurt swallow.

Ratings

Speech-language pathologists with experience in performing and interpreting FEES and videofluoroscopy were invited to rate the examinations. Twenty speech-language pathologists were approached and seventeen volunteered. All raters completed a questionnaire of videofluoroscopy and FEES experience and training.

The raters were instructed to score the clips using the Penetration-Aspiration Scale,¹³ viewing the clips no more than twice (see Appendix A). The Penetration-Aspiration Scale is an 8-point clinical scale for rating penetration and aspiration. The severity of the rating depends on the perceived depth of entry of food or fluid material into the airway and clearance of material from the airway. The scale has been evaluated for use with videofluoroscopy and FEES.^{13,14}

The raters were asked to rate the clips on CD1 first, then rate the clips on CD2 one week later without referring to their previous ratings. The raters were blinded to the participant details and to the pairing of the FEES and videofluoroscopy recordings. The ratings of two of the raters were excluded. One rater inadvertently swapped the disks, and the other's scores were illegible.

Assessment of Interrater and Intrarater Reliability

Intra- and interrater reliability were calculated separately for FEES and for videofluoroscopy using weighted Kappa on the original ratings. Kappa is a conservative measure (i.e., it tends to under-estimate agreement compared to other measures), and ranges from 0 (chance agreement) to 1 (complete agreement).

TABLE I.
Demographic Details of Study Participants.

Age	Sex	Diagnosis	Treatment History
78	F	Bilateral vocal-fold palsy	Tracheostomy
45	F	Suspected sarcoidosis	None
45	F	Cervical spine degeneration	Anterior cervical spine surgery
78	M	Cerebral small vessel disease	None
22	F	Congenital TOF*	Partial esophagectomy
56	M	Ewings sarcoma mandible	Partial mandibulectomy and glossectomy
64	M	Previous CVA† skull base tumor	None
58	M	Base of tongue carcinoma	Trans-oral laser resection, chemoradiotherapy
37	M	None (odynophagia)	None
58	F	Multiple sclerosis	Medication
56	M	Base of tongue carcinoma	Trans-oral laser resection, chemoradiotherapy
44	M	Suspected laryngo-pharyngeal reflux	None
74	M	Laryngeal carcinoma	Trans-oral laser resection
46	F	Laryngo-pharyngeal reflux	Medication
40	F	Systemic lupus erythematosus	Oesophageal dilatations

*TOF = tracheo-oesophageal fistula.
†CVA = cerebrovascular accident

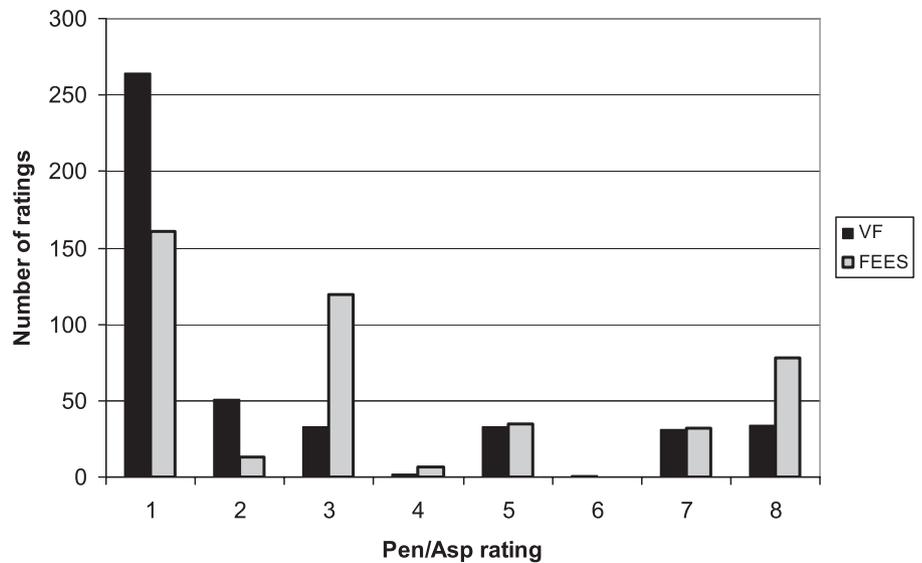


Fig. 1. The number of individual ratings made for penetration aspiration ratings from 1 to 8.

Assessment of Factors Affecting the Residue Rating

Five-way analysis of variance (ANOVA) was used to assess systematic differences in the ratings due to five factors: 1) examination (FEES or videofluoroscopy); 2) bolus type (liquid or yogurt); 3) rating (first or second); 4) rater (1–15); and 5) patient (1–15). Data were analyzed using SPSS for Windows (V12.0, SPSS Inc, Chicago, IL).

RESULTS

Scores

Fifteen raters twice scored penetration-aspiration from 30 recordings, yielding a total of 1,800 scores.

There was a wide range of Penetration-Aspiration Scale ratings from 1 to 8 although there were few ratings of 4 or 6 for both FEES and videofluoroscopy (see Fig. 1). This pattern is consistent with previous clinical Penetration-Aspiration Scale studies, including Rosenbek's description of the scale development and validation where there were few scores of 4 or 6.¹³

Raters

Fourteen raters returned the experience questionnaire. Eleven had attended a videofluoroscopy course and 12 had attended a FEES course. Clinical experience in FEES and videofluoroscopy was measured in examination-years, where one examination-year represents one examination performed and interpreted per week for one year (approximately 50 examinations). The mean (SD) experience was 6.0 (4.5) examination-years for videofluoroscopy; for FEES it was slightly lower at 4.9 (4.3) examination-years.

Assessment of Reliability

As would be expected, within-rater agreement was better than agreement between different raters, but both were subjectively good.¹⁵ Interrater reliability was 0.64 for FEES and 0.67 for videofluoroscopy. Intrarater reliability was 0.73 for FEES and 0.79 for videofluoroscopy.

Effect of Examination Type on Penetration-Aspiration Scale Scores

FEES scores were consistently higher than videofluoroscopy scores for both the liquid and yogurt boluses (see Fig. 2). That is, experienced clinicians perceived greater penetration and aspiration from FEES than from videofluoroscopy images of the same swallow.

Using ANOVA, this effect was assessed across all 15 raters and 30 boluses. The mean Penetration Aspiration Scale score from FEES recordings was more than 1 point higher than the mean score from videofluoroscopy recordings, statistically a highly significant effect (ANOVA, $F = 296$, $P < .001$; 95% CI for the difference, 1.01–1.27).

Effect of Other Factors on Penetration-Aspiration Scale Scores

Four of the five factors had statistically significant effects on score differences. The type of examination (FEES or videofluoroscopy) and differences between participants had the most significant impact on the scores (see Table II and Fig. 3). The high impact of participant differences on scores was unsurprising as the participants had a wide range of medical diagnoses, treatment histories, severity of dysphagia, and ages.

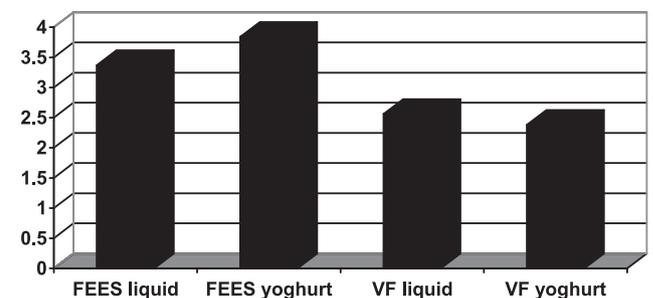


Fig. 2. Mean videofluoroscopy (VF) and fiberoptic endoscopic evaluation of swallowing (FEES) Penetration Aspiration Scale scores.

TABLE II.
Effects of Examination (VF vs. FEES) and Four Other Factors on Penetration Aspiration Scale Scores.

Factor	Mean Pen/Asp Score	P Value
Exam		
VF	2.47	<.001
FEES	3.61	
Rating		
First	3.04	.881
Second	3.03	
Bolus		
Liquid	2.97	.032
Yogurt	3.11	
Rater		
Lowest mean score	2.67	<.001
Highest mean score	3.98	
Subject		
Lowest mean score	1.16	<.001
Highest mean score	7.27	

For rater and subject, the overall range of scores is shown. P value is from ANOVA, and indicates the probability that the effect is a result of chance alone. FEES = fiberoptic endoscopic evaluation of swallowing; VF = videofluoroscopy; Pen/Asp = Penetration Aspiration.

Choice of Statistical Methods

An ANOVA model assumes data are normally distributed. When this is not the case, the statistical power of the test is reduced, and it becomes less likely to detect effects that are actually present. In our study, the principal effect we were interested in (i.e., the difference between examination types) was demonstrated at an extremely high level of significance ($P < 10^{-41}$). Nevertheless we also applied logistic regression, a more limited test suitable for non-normal data. The results were very similar to those from ANOVA; examination type was the most important effect, while bolus type had a weak effect on penetration aspiration scores and repeat assessment had no significant effect.

DISCUSSION

Key Findings of the Study

Our data clearly show that a rater's judgment of the severity of penetration or aspiration is affected by the type

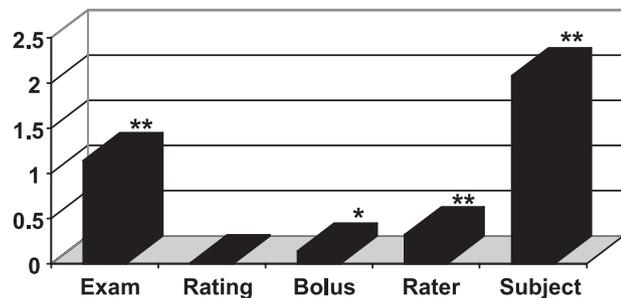


Fig. 3. Effects of different factors on Penetration Aspiration Scale scores. Effects are statistically significant at * $P < .05$ and ** $P < .001$. For rater and subject, the typical effect size is estimated by variance components analysis.

of examination image. The 15 raters consistently awarded higher (more severe) Penetration Aspiration Scale scores from FEES images than from videofluoroscopy images of the same swallows. This finding echoes those of our previous study showing that pharyngeal residue is perceived to be greater (more severe) from FEES than from videofluoroscopy.¹² In contrast to previous reports, it is clear that FEES and videofluoroscopy are not interchangeable for use in rating some parameters of disordered swallowing. Such use of these examinations in clinical practice may misrepresent a patient's swallowing function as improving or worsening, when any changes may be entirely attributable to a different examination method.

Whether the differences between FEES and videofluoroscopy have a clinically significant impact on predicting the likelihood of aspiration pneumonia requires further investigation.

Rating Scale

The Penetration-Aspiration Scale is a sensitive tool for rating precise degrees of entry of material into the airway, the sensorimotor response, and the effectiveness of any spontaneous attempts to clear material from the airway. The scale is an important clinical tool and has provided some much needed standardization of FEES and videofluoroscopy scoring. The challenge of achieving consistent and therefore reliable scoring of dysphagia parameters from videofluoroscopy is well recognized.¹⁶ FEES may be reliable for assessing penetration and aspiration using the Penetration-Aspiration Scale,¹⁴ but whether it is a reliable technique for the assessment of a wider range of swallowing and dysphagia parameters remains unknown. Given the subjective and variable methods of conducting and interpreting FEES and videofluoroscopy, neither examination provides the "right answer" and so can not be considered the gold standard for swallowing examination.

Strengths of the Study

Number of Raters

Fifteen independent raters from 12 centers across the United Kingdom rated the examinations. Previous studies comparing FEES and videofluoroscopy have used few or an unspecified number of raters.^{3,7-11} Using a large number of raters reduces the effects of random scoring errors or bias in individual raters.

Validated Rating Scale for Both Types of Exams

This is the first study to rate the severity of penetration and aspiration from simultaneous FEES and videofluoroscopy using a scale validated for both examinations.^{13,14} Previous studies comparing FEES and videofluoroscopy have scored penetration and aspiration as present or absent which limits the clinical relevance of their findings.

Limitations

Number of Swallows Rated

Each rater scored 120 swallows across two sessions, with each rating session taking approximately 90 minutes

to complete. In a prestudy pilot, this was the maximum time the raters felt able to rate swallows without losing concentration. Increasing the number of swallows would have resulted in rater drop out.

Number and Range of Boluses

Increasing the number of boluses and the range of food and fluid consistencies would have provided additional useful information. For the purposes of this study, all participants were required to swallow identical boluses to ensure that raters could not identify the paired recordings. We recruited participants with dysphagia ranging from mild to severe to ensure a wide range of ratings across the Penetration Aspiration Scale. Increasing the number of boluses and range of consistencies would have placed the participants with severe dysphagia at unacceptable risk of severe aspiration.

CONCLUSION

Penetration and aspiration are perceived to be more severe from FEES than from videofluoroscopy. This has serious implications for the interchangeable use of these examinations in clinical practice. Further research looking at the impact of penetration-aspiration scoring on clinical recommendations and health outcomes is required.

BIBLIOGRAPHY

1. Kidder TM, Langmore SE, Martin BJ. Indications and techniques of endoscopy in evaluation of cervical dysphagia: comparison with radiographic techniques. [Review] *Dysphagia* 1994;9:256–261.
2. Langmore SE. Evaluation of oropharyngeal dysphagia: which diagnostic tool is superior? [Review] *Curr Opin Otolaryngol Head Neck Surg* 2003;11:485–489.
3. Logemann JA, Rademaker AW, Pauloski BR, Ohmae Y, Kahrilas PJ. Normal swallowing physiology as viewed by videofluoroscopy and videoendoscopy. *Folia Phoniatrica et Logopedica* 1998;50:311–319.
4. Aviv JE. Prospective, randomized outcome study of endoscopy versus modified barium swallow in patients with dysphagia. *Laryngoscope* 2000;110:563–574.
5. Doggett DL, Tappe KA, Mitchell MD, Chapell R, Coates V, Turkelson CM. Prevention of pneumonia in elderly stroke patients by systematic diagnosis and treatment of dysphagia: an evidence-based comprehensive analysis of the literature. [Review] *Dysphagia* 2001;16:279–295.
6. Perry L, Love CP. Screening for dysphagia and aspiration in acute stroke: a systematic review. [Review] *Dysphagia* 2001;16:7–18.
7. Langmore SE, Schatz K, Olson N. Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. *Ann Otol Rhinol Laryngol* 1991;100:678–681.
8. Madden C, Fenton J, Hughes J, Timon C. Comparison between videofluoroscopy and milk-swallow endoscopy in the assessment of swallowing function. *Clin Otolaryngol Allied Sci* 2000;25:504–506.
9. Perie S, Laccourreye L, Flahault A, Hazebrucq V, Chaussade S, St Guily JL. Role of videoendoscopy in assessment of pharyngeal function in oropharyngeal dysphagia: comparison with videofluoroscopy and manometry. *Laryngoscope* 1998;108:1712–1716.
10. Rao N, Brady SL, Chaudhuri G, Donzelli JJ, Wesling MW. Gold-standard? Analysis of the videofluoroscopic and fiberoptic endoscopic swallow examinations. *J Applied Res* 2003;3:89–96.
11. Wu CH, Hsiao TY, Chen JC, Chang YC, Lee SY. Evaluation of swallowing safety with fiberoptic endoscope: comparison with videofluoroscopic technique. *Laryngoscope* 1997;107:396–401.
12. Kelly AM, Leslie P, Beale T, Payten C, Drinnan MJ. Fiberoptic endoscopic evaluation of swallowing and videofluoroscopy: does examination type influence perception of pharyngeal residue severity? *Clin Otolaryngol* 2006;31:425–432.
13. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia* 1996;11:93–98.
14. Colodny N. Interjudge and intrajudge reliabilities in fiberoptic endoscopic evaluation of swallowing (fees) using the penetration-aspiration scale: a replication study. *Dysphagia* 2002;17:308–315.
15. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977;33:159–174.
16. Stoeckli SJ, Huisman TA, Seifert B, Martin-Harris BJ. Interrater reliability of videofluoroscopic swallow evaluation. *Dysphagia* 2003;18:53–57.

APPENDIX A

The Penetration-Aspiration Scale

1. Material does not enter the airway.
2. Material enters the airway, remains above the vocal folds, and is ejected from the airway.
3. Material enters the airway, remains above the vocal folds, and is not ejected from the 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 from the trachea despite effort.
8. Material enters the airway, passes below the vocal folds, and no effort is made to eject.