

연하 장애의 정확한 진단을 위한 굴곡형 내시경 검사 항목의 표준화를 위한 연구

김보영¹ · 이진² · 김보혜³ · 박하나로⁴ · 박성준⁵ · 송창면⁶ · 정은재⁷ · 권택균⁷ · 진영주²

¹인제대학교 의과대학 상계백병원 이비인후과, ²원광대학교 의과대학 원광대병원 이비인후과, ³동국대학교 의과대학 동국대 일산병원 이비인후과, ⁴성균관대학교 의과대학 삼성창원병원 이비인후과, ⁵중앙대학교 의과대학 이비인후과, ⁶한양대학교 의과대학 이비인후과, ⁷서울대학교 의과대학 이비인후과

Standardization of FEES Evaluation for the Accurate Diagnosis of Dysphagia

Bo Young Kim, M.D., Ph.D.¹, Jin Lee, M.D.², Bo Hae Kim, M.D., Ph.D.³, Hanaro Park, M.D.⁴, Sung Joon Park, M.D.⁵, Chang Myeon Song, M.D., Ph.D.⁶, Eun-Jae Chung, M.D., Ph.D.⁷, Tack-Kyun Kwon, M.D., Ph.D.⁷, Young Ju Jin, M.D., Ph.D.²

¹Department of Otorhinolaryngology-Head and Neck Surgery, Sanggye Paik Hospital, College of Medicine, Inje University, Seoul, ²Department of Otorhinolaryngology-Head & Neck Surgery, Wonkwang University Hospital, Wonkwang University College of Medicine, Iksan, ³Department of Otorhinolaryngology-Head and Neck Surgery, Dongguk University Ilsan Hospital, Dongguk University College of Medicine, Goyang, ⁴Department of Otorhinolaryngology-Head and Neck Surgery, Samsung Changwon Hospital, Sungkyunkwan University School of Medicine, Changwon, ⁵Department of Otorhinolaryngology-Head and Neck Surgery, College of Medicine, Chung-Ang University, Seoul, ⁶Department of Otolaryngology-Head and Neck Surgery, Hanyang University College of Medicine, Seoul, ⁷Department of Otorhinolaryngology, College of Medicine, Seoul National University, Seoul, Korea

Objective: Fiberoptic endoscopic evaluation of swallowing (FEES) is a standard diagnostic tool for swallowing disorders. However, it has not been used frequently in Korea because of the long test time, low cost, and the absence of a standard evaluation system. The purpose of this study was to suggest a standard fill-out form for the FEES result.

Methods: From February 2019 to June 2020, a total of 98 FEES tests were performed by an otolaryngologist (JYJ) at the Wonkwang University Hospital. After the exclusion of 68 cases, 30 cases were analyzed twice by 4 raters with over 5 years of experience as otolaryngologists working in various hospitals. The results were measured for the rater's test-retest reliability and inter-rater consistency.

Results: Cohen's kappa values for measuring the intra-rater consistency of the four raters were 0.984, 0.887, 0.848, and 0.930, respectively, meaning very good alignment of 0.8 or more, respectively. The Fleiss Kappa value for measuring inter-rater consistency was 0.276, meaning 'fair' for values of 0.2 or more. To examine consistency, an intraclass correlation coefficient (ICC) analysis conducted by assuming the grading score to be a constant continuous variable gave an ICC value of 0.729 ($P < 0.001$), showing a very reliable tendency.

Conclusion: In this study, all the items of the fill-out form were rated using a three-step grading scale, so the degree of agreement was high when performed twice by the same rater, but the degree of agreement among raters was relatively low. Therefore, our fill-out form for FEES will be useful in evaluating the improvement of a patient over the course of clinical treatment. (JKDS 2022;12:59-63)

Keywords: Fiberoptic endoscopic evaluation of swallowing, Swallowing disorder, Fill out form

투고일: 2021년 9월 29일, 심사일: 2021년 9월 30일, 게재확정일: 2021년 12월 2일

책임저자: 진영주, 전라북도 익산시 무왕로 895

(54538) 원광대학교 의과대학 원광대학교병원 이비인후과학교실

Tel: 063) 859-1445, Fax: 063) 859-1440, E-mail: chindol1@wku.ac.kr



This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License, which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyrights © The Korean Dysphagia Society, 2022.

서론

연하 장애는 음식을 삼키는 과정에 어려움이 발생하는 질병으로 주로 뇌졸중, 중추 및 말초 신경의 퇴행성 질환, 외상성 뇌질환, 두경부 종양, 근육 질환 환자들에게 동반된다. 뇌졸중 환자의 8.2-80%, 파킨슨병 환자의 11-81%, 두경부 외상 환자의 27-30%, 지역사회 획득 폐렴 진단된 고령 환자의 91.7%에서 연하 장애가 보고되었는데, 악화 시 기도 흡인, 탈수, 영양실조 등이 유발되고 사망까지 이를 수 있어, 정확한 진단과 신속한 치료는 환자들의 삶의 질 향상에 매우 중요하다¹⁻³.

연하장애 진단을 위해 가장 유용한 진단적 검사는 비디오 투시 연하검사(Video Fluoroscopic Swallowing Study, VFSS)와 내시경 연하검사(Fiberoptic endoscopic evaluation of swallowing, FEES)이다. 비디오투시검사는 투시조영(fluoroscopy)을 사용하여 삼킬 시 발생하는 투과(penetration), 흡인(aspiration), 비정상적 인두 저류(abnormal pharyngeal residue)를 감지하여, 연하곤란을 유발하는 기능적 구조적 원인을 파악할 수 있다⁴. 하지만 앉은 자세를 취할 수 있는 환자들에게만 검사가 가능하고, 별도의 투시조영 장비가 필요하며, 검사시마다 환자가 방사능에 노출된다는 단점이 있다. 한편, 1988년에 Langmore에 의해 처음 보고된 내시경 연하검사는, 굴곡형 내시경을 코에 넣어, 환자가 음식물을 삼킬 때 직접 인후두를 관찰하는 방법으로 다음과 같은 장점이 있다⁵. 첫째 X-ray를 찍어 관찰하는 비디오투시검사와는 달리 직접적으로 내시경을 넣어 인후두 구조물을 정확히 관찰할 수 있다. 두번째, 삼키는 능력 및 과도한 인후두내 분비물에 대한 환자의 반응을 직접 눈으로 보고 평가할 수 있다. 셋째, 내시경을 이용해 피열연골(arytenoid), 후두개, 가성대를 직접 접촉하거나 공기 자극을 주어, 후두내전반사(Laryngeal adductor reflex)가 정상적으로 발생하는지 시험해 봄으로써, 인후두의 감각기능을 직접적으로 검사해 볼 수 있다. 넷째, 연하재활운동시 굴곡형 내시경으로 검사를 직접 환자에게 보여줌으로써, 환자 교육을 위한 생체 되먹임 도구로 사용될 수 있다. 다섯번째, 검사시마다 방사능 노출이 없으며, 환자의 자세나 장소에 상관없이 검사 가능하다^{5,6}. 이와 같은 이유로, 비디오투시검사와 함께 내시경 연하검사는 연하 장애의 정확한 진단을 위해 매우 안전하고 믿을 만한 검사 방법이며, 흡인으로 인한 폐렴의 유발률을 분석한 연구들에서도 두 방법 사이에 진단차이가 없는 것으로 보고되어서, 상황에 따라 상호 보완적으로 사용 되어져야 한다^{1,7}.

하지만 이처럼 간편하고 많은 정보를 주는 검사임에도

불구하고, 내시경연하검사는 검사 소요시간에 비해 낮은 수가, 오랜 검사시간, 검사 방법 및 다양한 평가체계 등의 문제로, 국내에서 활발히 시행되지 있지 않으며, 특히 검사에 대한 표준 양식 및 방법이 정해져 있지 않아서, 대부분의 병원에서는 비디오투시검사 결과지를 갈음해 사용하고 있다. 본 연구에서는 연하 장애의 정확하며 보편적인 진단을 위해 FEES 검사 항목 및 분석 체계의 표준화를 위한 양식을 제시하고자 한다.

대상 및 방법

2019년 2월부터 2020년 6월까지 원광대 병원 이비인후과 전문의(JY)가 시행한 총 98례의 굴곡형 내시경 연하 검사 결과 중 화질이 너무 나빠서 판독하기 어렵거나, 인두 저류를 측정하지 않았거나, 검사 시작 전에 이미 투과 흡인이 발생하고 있는 68명을 제외하고, 30례의 영상을 선별하였다.

서로 다른 종합병원에 근무하고 있는 이비인후과 전문의 경력 5년 이상된 총 4명의 평가자가 2차례 영상을 분석 하였으며, 1차 분석과 2차 분석 사이에는 최소한 24시간 이상의 시간차를 두도록 하였다. 분석 결과에 대해 평가자내 일치도와 평가자간 일치도를 측정하였다. 내시경 연하 검사 결과 분석 양식은 여러 이전 논문을 참고하여, 간단한 형식으로 작성하였다.(Fig. 1)

1. 통계분석

IBM SPSS statistics version 26 (IBM Corp., Armonk, N.Y., USA)을 사용하여 개별 평가자의 검사-재검사 일치도(test-retest reliability)는 Cohen's kappa를 계산하여 측정하였고 평가자 간의 일치도(inter-rater consistency)는 Kleiss kappa를 사용하여 평가하였다. 평가자간 측정의 연관성(correlation)을 평가하기 위하여 급내상관계수(intraclass correlation coefficient, ICC)를 사용하였다. 급내상관계수는 연속 변수의 분석에 사용하지만 내시경연하검사 grading을 등 간격연속변수로 가정하고 분석을 시행하였다.

결과

총 30례의 비디오 이미지 중, 10례는 두경부 암 환자(후두암 6명, 하인두암 3명, 편도암 1명)였으며 이들의 평균 연령은 69.4세였고, 모두 남성이었다. 20례는 연하곤란을 주소로 타과에서 의뢰된 환자였으며, 남녀 비는 13:7이며, 평균 연령은 67.1세였다.

FEES report					
Tongue movement			○ Normal ○ Impaired		
Mouth closure			○ Normal ○ Impaired		
Mouth opening			○ Normal ○ Impaired		
Gag reflex			○ Normal ○ Impaired		
Vocal fold movement			○ Normal ○ Impaired (Rt, Lt, Both)		
Without food			With food		
Categories	Vallecular pooling	Pyriform sinus pooling	Vallecular pooling	Pyriform sinus pooling	Laryngeal penetration /tracheal aspiration
Rating	0=no pooling 1≤50% 2>50%	0=no pooling 1≤50% 2>50%	0=no pooling 1≤50% 2>50%	0=no pooling 1≤50% 2>50%	0=no laryngeal penetration 1=laryngeal penetration 2=tracheal aspiration
Milk					
Yogurt					
Recommendation					

Fig. 1. FEES report form.

Table 1. Individual intra-rater consistency.

	Rater1	Rater2	Rater3	Rater4
Cohen's kappa	0.984	0.887	0.848	0.930
SE	0.016	0.037	0.037	0.026
P value	<0.001	<0.001	<0.001	<0.001

4명 개별 평가자의 검사-재검사 일치도(intra-rater consistency)를 측정하기 위한 Cohen's kappa 값은 각각 0.984, 0.887, 0.848, 0.930으로 모두 0.8 이상의 very good agreement를 보였다.(Table 1)

4명 평가자 간의 검사결과 일치도(inter-rater consistency)를 측정하기 위한 Fleiss Kappa 값은 0.276으로 측정되었으며, 0.2 이상으로 'fair'에 해당되었다. 일치도(Consistency)의 경향성을 보기 위하여 grading 점수를 등간격 연속변수로 가정하고 시행한 ICC 분석에서 ICC값은 0.729, $P<0.001$ 으로 매우 신뢰할 만한 경향성을 보인다.(Table 2)

고찰

본 연구에서 제시한 내시경연하검사 양식을 사용하여 평가한 인두 저류(residue), 투과(penetration), 흡인(aspiration)에 대한 서로 다른 평가자 4명의 개별 평가자 내 일치도(intra-rater consistency)는 0.8 이상으로 very good agreement를 보였으며, 4명의 평가자 간 일치도는 Fleiss kappa값이 0.276으로 Fair에 해당하지만 높은 편은 아니었다. 일치도(Consistency)의 경향성을 보기 위하여 grading 점수를 등간격 연속변수로 가정하고 시행한 ICC 분석에서 ICC값은 0.729, $P<0.001$ 으로 이는 검사자 간에 정확히 같은 점수를 주는 일치도(consistency)는 아주 높지 않았으나, 검사자들이 준 점수의 경향성(tendency)은 유의미하게 높음을 의미하였다. 본 연구에서 제시한 FEES 양식은 서로 다른 검사자가 시행하였을 때 같은 등급으로 평가하는 평가자 간 일치도(inter-rater consistency)는 비교적 낮았으나, 모든 항목이 비교적 간단한 3단계 채점(grading)을 채택함으로써 같은 검사자가 여러 번 시행하였을 때 일치도는 높았다. 특히 FEES가 널리 행해지지 않고 있는 현실을 고려했을 때, 평가자의 FEES 경험이 적었던 것이 평가자간 일치도가 낮

Table 2. Inter-rater consistency.

	Statistics	Value	95% CI	P value
Absolute agreement	Fleiss Kappa	0.276	0.275-0.278	<0.001
Correlation consistency	ICC*	0.729	0.651-0.793	<0.001

*Intraclass Correlation Coefficient, assuming the data is an ordinal scale.

있던 이유가 될 수 있으며, 다만 개별 평가자의 검사-재검사 일치도 매우 높아, 제한된 환경 특히 한 환자의 임상 경과에 따른 호전 정도를 비교할 때에 유용할 것으로 생각한다. 또한 평가자내 일치도는 높으나 평가자간 일치도가 낮은 것은 추후 FEES 교육을 통해 해결할 수 있는 문제임을 보여주고 있다.

본 연구는, FEES 검사에 맞는 적절한 검사 양식을 제시하고 자 시행되었다. 비디오투시검사의 scoring 체계는 1996년 Rosenbeck 등이 제시한 이전의 9단계를 수정하여 삼킴 과정 중 발생하는 흡인과 투과외의 정도 등에 따라 8단계로 구분한 Penetration-Aspiration Scale (PAS)이 가장 널리 사용되고 있다⁸. PAS 1은 기도로 음식물이 들어가지 않을 때, PAS 2, 3은 성대 위에 머무를 때, PAS 4, 5는 음식이 성대를 접촉했을 때, PAS 6은 음식이 성대 아래로 내려가 기관지도 들어갔으나 다시 나올 때, PAS 7은 음식이 성대 아래로 들어갔고, 기침 등의 노력에도 기관지 밖으로 나오지 않을 때, PAS 8은 성대 아래로 들어갔고, 기침 등의 노력도 없을 때의 상황으로 나뉘어져 있는데, 이는 비디오투시 검사에서 보이는 검사 결과의 판독에 용이하게 구분되어 있다. 하지만 내시경 연하검사시엔, 환자가 음식물을 삼키는 순간엔 화면에 아무것도 보이지 않게 되어(white out), 음식물이 성대에 접촉하는지, 아래로 내려가는 지에 대해 판단이 어려우며, 다만 삼킨 전 흡인과 삼킨 후 흡인 여부, 기도 밖으로 내보내려는 의지가 있는지 정도가 판단이 가능하여 PAS score와는 맞지 않다. 특히 내시경 연하검사는 내시경으로 직접 보는 검사로, 음식물 투여 전후 인두 저류에 대한 관찰이 가능하며, 인두 저류가 검사 전부터 관측될 때에는 어느 정도 삼킴 기능 저하, 흡인 및 투과를 예측할 수 있는데 이것의 검사를 용이하게 할 수 있는 장점이 있다. 따라서 내시경 연하검사 검사에 적합한 scoring 체계의 필요성이 부각되었다. 한편, 2016년 W.Pilz 등은 내시경 검사 해석을 위해 다른 관찰자 2명의 결과를 분석하였으며, 분석을 위해서 우리 연구에서 사용한 것과 같은 순위 척도(ordinal rating scale)를 사용하였는데 삼킴 후 인두 저류 정도와 흡인 및 투과 정도를 비교하였을 때, 관찰자의 내시경 연하검사 결과 해석에 대한 일치율은 연하 장애의 원인

질병, 관찰자에 따른 차이에 의해 결정되는 것이 아니라 주로, 검사를 위해 투여한 액체의 점도 차이(thick liquid, thick liquid)에 의해 영향 받는 것으로 결론 내렸다⁹. 따라서 우리가 제시한 내시경 연하 검사 양식에서는 요플레와 우유와 같이 점도가 다른 두가지 음식물을 투여해 검사를 시행하였다. 한편 Baijens, Laura WJ 등은 84명의 연하곤란 환자들에 대해 역시 이전 다른 검사들에서처럼 내시경연하검사를 위해 10 cc thick liquid를 10회 삼키게 한 후, 흡인의 유무로 결정하였는데, 단지 3-4회의 삼킴 검사 시도한 경우엔 검사결과가 저 평가될 수 있었으며, 삼킴 검사 횟수가 증가할수록 민감도가 증가하여, 정확한 검사 결과는 삼키기 검사 횟수에 크게 의존한다고 보고하였다¹⁰.

삼킴 전후에 발견되는 인두 저류는 식사 후 흡인이 발생할 수 있는 중요한 예측인자이다¹¹. 인두 저류는 주로 후두개곡(valleculae)이나, 양측 이상와(Pharyngeal residue)에 발생하는데, 저류의 심각성(severity) 정도를 객관적으로 정확히 기술하는 것은 어려워서, 다양한 종류의 분석방법들이 보고되었다. 인두 저류의 유무¹², 저류된 양에 따른 순위 척도(ordinal score)¹³⁻¹⁵, 검사를 위해 투입한 양(bolus)에 비해 비디오투시검사시 x ray 상으로 관찰되는 인두 저류 비율의 추정¹⁶, 컴퓨터 이미지 분석을 통한 정량화¹⁷ 등의 방법으로서의 측정이 보고되었는데, Neubauer 등은 systemic review를 시행하였으며, 최종 선발된 7개의 인두 저류 척도 관련 논문 중, 예일 인두 저류 척도(Yale pharyngeal residue severity rating scale)가 가장 유효하고 신뢰할 만한 척도라고 보고하였다^{15,18}.

본 연구의 장점은 첫째, 내시경 연하 검사가 능숙한 1인의 동일한 이비인후과 전문의에 의해 이뤄졌다. 둘째, 검사 결과 분석은 매우 독립된 4개의 센터에서 각각 이뤄졌으며 각각의 측정자는 24시간의 시간차를 두고 2회 측정하였다. 셋째, 국내에서 처음으로 내시경 연하 검사에 대한 표준 검사 양식을 만들었으며, 여러 독립된 측정자들을 통해 재현성 있고 믿을만한 검사지임을 증명하였다. 본 연구의 단점은 다음과 같다. 첫째, 검사시 투입하는 음식물의 양을 일정하게 하지 않았다. 둘째, 검사 횟수가 증가할수록 연하 장애가 저평가 되는 것이 감소한다고 보고 되었었는데, 본 연구

의 검사양식에서는 그 점에 대한 고려를 하지 않았다. 셋째, 측정자는 내시경 연하 검사 분석에 대해 충분한 교육을 시행하지 않고 검사가 이뤄졌다. 하지만 그럼에도 검사자들간 일치도가 높게 나와 초보자도 쉽게 사용할 수 있는 믿음만한 검사 툴이라고 할 수 있다. 넷째, 연하장애 평가를 위해 가장 많이 사용되는 비디오 투시 연하검사 결과와의 비교 연구를 하지 않았으며 대상 case가 30례로 비교적 작은수의 영상을 분석하였다.

결론

본 연구에서 제시한 내시경 연하검사 양식은 모든 항목이 3 단계 척도(grading)를 채택함으로써 같은 검사자가 여러 번 시행하였을 때 일치도는 높았으나 서로 다른 검사자가 시행하였을 때 같은 척도로 평가하는 검사자간 일치도는 비교적 낮았다. 검사자의 평가의 경향성은 매우 높아 제한된 환경, 특히 한 환자의 임상 경과에 따른 호전 정도를 비교할 때에 유용할 것으로 생각한다.

ACKNOWLEDGEMENTS

이 연구는 ‘2019년도 대한 연하장애학회 연구비’ 지원을 받아 작성하였습니다.

REFERENCES

1. Aviv JE. Prospective, randomized outcome study of endoscopy versus modified barium swallow in patients with dysphagia. *The Laryngoscope*. 2000;110:563-74.
2. Wilkins T, Gillies RA, Thomas AM, Wagner PJ. The prevalence of dysphagia in primary care patients: a HamesNet Research Network study. *J Am Board Fam Med*. 2007;20:144-50.
3. Takizawa C, Gemmell E, Kenworthy J, Speyer R. A systematic review of the prevalence of oropharyngeal dysphagia in stroke, Parkinson's disease, Alzheimer's disease, head injury, and pneumonia. *Dysphagia*. 2016;31:434-41.
4. Kim JY, Koh ES, Kim HR, Chun SM, Lee SU, Jung SH. The Diagnostic Usefulness of the Fiberoptic Endoscopic Evaluation of Swallowing. *Journal of the Korean Academy of Rehabilitation Medicine*. 2011;35:14-22.
5. Langmore SE, Kenneth SM, Olsen N. Fiberoptic endoscopic examination of swallowing safety: a new procedure. *Dysphagia*. 1988;2:216-9.
6. Langmore SE. History of fiberoptic endoscopic evaluation of swallowing for evaluation and management of pharyngeal dysphagia: changes over the years. *Dysphagia*. 2017;32:27-38.
7. Langmore SE. Evaluation of oropharyngeal dysphagia: which diagnostic tool is superior? *Current opinion in otolaryngology & head and neck surgery*. 2003;11:485-9.
8. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996;11:93-8.
9. Pilz W, Vanbelle S, Kremer B, van Hooren MR, van Becelaere T, Roodenburg N, et al. Observers' agreement on measurements in fiberoptic endoscopic evaluation of swallowing. *Dysphagia*. 2016;31:180-7.
10. Baijens LW, Speyer R, Pilz W, Roodenburg N. FEES protocol derived estimates of sensitivity: aspiration in dysphagic patients. *Dysphagia*. 2014;29:583-90.
11. Murray J, Langmore SE, Ginsberg S, Dostie A. The significance of accumulated oropharyngeal secretions and swallowing frequency in predicting aspiration. *Dysphagia*. 1996;11:99-103.
12. Dejaeger E, Pelemans W, Ponette E, Joosten E. Mechanisms involved in postdeglutition retention in the elderly. *Dysphagia*. 1997;12:63-7.
13. Farneti D. Pooling score: an endoscopic model for evaluating severity of dysphagia. *Acta Otorhinolaryngologica Italica*. 2008;28:135.
14. Kaneoka AS, Langmore SE, Krisciunas GP, Field K, Scheel R, McNally E, et al. The Boston residue and clearance scale: preliminary reliability and validity testing. *Folia Phoniatrica et Logopaedica*. 2013;65:312-7.
15. Neubauer PD, Rademaker AW, Leder SB. The Yale Pharyngeal Residue Severity Rating Scale: an anatomically defined and image-based tool. *Dysphagia*. 2015;30:521-8.
16. Eisenhuber E, Schima W, Schober E, Pokieser P, Stadler A, Scharitzer M, et al. Videofluoroscopic assessment of patients with dysphagia: pharyngeal retention is a predictive factor for aspiration. *American Journal of Roentgenology*. 2002;178:393-8.
17. 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.
18. Neubauer PD, Hersey DP, Leder SB. Pharyngeal residue severity rating scales based on fiberoptic endoscopic evaluation of swallowing: a systematic review. *Dysphagia*. 2016;31:352-9.