

Inter-rater Reliability of Videofluoroscopic Dysphagia Scale

Dae Ha Kim, M.D., Kyoung Hyo Choi, M.D., Ph.D., Hong Min Kim, M.D., Jung Hoi Koo, M.D.¹,
Bo Ryun Kim, M.D.², Tae Woo Kim, M.D.³, Joo Seok Ryu, M.D.⁴, Sun Im, M.D.⁵, In Sung Choi, M.D.⁶,
Sung Bom Pyun, M.D.⁷, Jin Woo Park, M.D.⁸, Jin Young Kang, M.D.⁹, Hee Seung Yang, M.D.¹⁰

Department of Rehabilitation Medicine, Asan Medical Center, Seoul 138-042, ¹Department of Rehabilitation Medicine, Kangneung Asan Hospital University of Ulsan College of Medicine, Kangneung 211-711, ²Department of Rehabilitation Medicine, University of Jeju College of Medicine, Jeju 690-767, ³Department of Rehabilitation Medicine, Kangneung Dong-in Hospital, Kangneung 210-111, ⁴Department of Rehabilitation Medicine, CHA University College of Medicine, Seongnam 463-712, ⁵Department of Rehabilitation Medicine Catholic University of Korea, College of Medicine, Bucheon 420-717, ⁶Department of Rehabilitation Medicine, Chonnam National University Medical School, Gwangju 501-757, ⁷Department of Physical Medicine and Rehabilitation, Korea University College of Medicine, Seoul 136-705, ⁸Department of Physical Medicine and Rehabilitation, Dongguk University College of Medicine, Goyang 410-773, ⁹Department of Physical Medicine and Rehabilitation, Won-Gwang University College of Medicine, Gunpo 435-040, ¹⁰Department of Physical Medicine and Rehabilitation, Seoul Veteran's Hospital, Seoul 134-791, Korea

Objective To investigate the inter-rater agreement using the Videofluoroscopic Dysphagia Scale (VDS).

Method The present study was designed as a multicenter, single-blind trial. A Videofluoroscopic Swallowing Study (VFSS) was performed using the protocol described by J.A Logemann. Thick-fluid, pureed food, mechanically altered food, regularly textured food, and thin-fluid boluses were sequentially swallowed. Each participant received a 3 ml bolus followed by a 5 ml bolus of each food material, in the order mentioned above. All study procedures were video recorded. Discs containing these video recordings in random order were distributed to interpreters who were blinded to the participant information. The video recordings were evaluated using a standardized VDS sheet and the inter-rater reliability was calculated.

Results In total, 100 patients participated in this study and 10 interpreters analyzed the findings. Inter-rater reliability was fair in terms of lip closure (κ : 0.325), oral transit time (0.253), delayed triggering of pharyngeal swallowing (0.300), vallecular residue (0.275), laryngeal elevation (0.345), pyriform sinus residue (0.310), coating of the pharyngeal wall (0.310), and aspiration (0.393). However, other parameters of the oral phase were lower than those of the pharyngeal phase (0.06-0.153). Moreover, the summation of VDS reliability (intraclass correlation coefficient: 0.556) showed moderate agreement.

Conclusion VDS shows a moderate rate of agreement for evaluating the swallowing function. However, many of the parameters demonstrated a lower rate of agreement, particularly the oral phase parameters.

Key Words VDS, Reliability, Inter-rater, Dysphagia, VFSS

Received June 29, 2012; Accepted August 16, 2012

Corresponding author: Kyoung Hyo Choi

Department of Rehabilitation Medicine, Asan Medical Center, 388-1, Poongnap-dong, Songpa-gu, Seoul 138-042, Korea

Tel: +82-2-3010-3800, Fax: +82-2-3010-6964, E-mail: khchoi@amc.seoul.kr

© This is an open-access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

Copyright © 2012 by Korean Academy of Rehabilitation Medicine

INTRODUCTION

Dysphagia is a frequent result of a stroke, brain tumor, or neurodegenerative disease. Many authors have tried to detect swallowing abnormalities (particularly aspiration) using non-radiographic observations, yet, these methods demonstrate poor sensitivity and specificity.^{1,2} The Videofluoroscopic Swallowing Study (VFSS) has been the gold standard for evaluating patients with swallowing disorders for many years.^{3,4} VFSS can detect oral, pharyngeal, and esophageal dysphagia; however, it demonstrates a limited ability in predicting the prognosis of dysphagia. Among many recent attempts to quantify and predict the prognosis of dysphagia, the Functional Dysphagia Scale (FDS), as reported by Han et al.⁵ in 2001, is a useful tool, correlating well with the ASHA-NOMS (American Speech-Language-Hearing Association National Outcomes Measurement System).⁶ However, despite its value in explaining the severity of dysphagia, FDS does not predict the long-term prognosis of dysphagia, which is important due to the close relationship between prolonged dysphagia, lower respiratory tract infection, and high mortality.^{7,8}

The Videofluoroscopic Dysphagia Scale (VDS) can be used to predict the long-term prognosis of dysphagia patients following stroke. Han et al.⁹ define the long-term prognosis of dysphagia based on the occurrence of any aspiration/penetration event after 6 months from the onset of dysphagia. VDS consists of 14 items with weighted values, and also shows good correlation with aspiration/penetration occurring 6 months after the initial onset of dysphagia. The 14 items in VDS (Appendix 1) represent oral (lip closure, bolus formation, mastication, apraxia, premature bolus loss, and oral transit time) and pharyngeal (pharyngeal triggering, vallecular and pyriform sinus residues, laryngeal elevation and epiglottic closure, pharyngeal coating, pharyngeal transit time, and aspiration) functions that can be observed by VFSS. VDS can also express the severity of dysphagia in a quantifiable score; however, limitations regarding the subjectivity of its results have been noted in previous studies. Stoeckli et al.¹⁰ report high interobserver reliability for some of the parameters used to evaluate aspiration and penetration, but low reliability for other oral and pharyngeal phase parameters. Although their study did not evaluate VDS, it suggests that the results of VFSS can be subjective on several parameters. Since VDS is measured based on the

findings of VFSS, the results may also be dependent on the observer; furthermore, there have not been any studies on its inter-rater reliability of VDS. Therefore, in this study, we investigate the inter-rater reliability of VDS.

MATERIALS AND METHODS

Participants

This study was designed as a multicenter (10 rehabilitation centers), single-blind trial. Patients who exhibited any symptoms of difficulty in swallowing were recruited. The criteria for inclusion were patients with (1) a history of aspiration symptoms, such as coughing or choking; (2) symptoms clinically suspicious of dysphagia, such as reduced gag reflex or delayed swallowing reflex; and (3) a history of the use of an alternative feeding method, such as a nasogastric tube. Patients who could not sit or those who had difficulty maintaining consciousness were excluded. All of the recruited patients who agreed to participate in our study underwent VFSS from January through June in 2011. The protocol for this study was approved by the Institutional Review Board of Seoul Asan Hospital.

VFSS protocol

VFSS was conducted by two physiatrists using fluoroscopy. The first physiatrist was a professor with 15 years of experience with VFSS; the second physiatrist was a resident physician. A modified version of the protocol used in Logemann's study¹¹ was used. The protocol consisted of 2 trials. The first trial was performed with the fluoroscopy projected from the lateral side of patient. Patients were asked to sit on a chair and then turn 90 degrees away from the fluoroscopy in order to form the lateral projection position. Patients were given 3 and 5 ml boluses of a thick-fluid mixture that contained barium (the viscosity was above 1750 centipoise (cP) using a syringe, followed by 3 and 5 ml boluses of a pureed diet, mechanically altered diet, and regularly textured food using a spoon. All of the food samples were administered two times. The last step of the first trial consisted of 3 ml and 5 ml boluses of thin-fluid mixture with barium (the viscosity was 1-50 cP) that was administered using a syringe; finally, 2 drinks of a thin-fluid mixture was administered using a cup.¹² The second trial was performed as an anterior-posterior projection with the patient sitting in an upright position. During the second trial, patients were asked to drink a thin-fluid mixture from a cup. If there was a large

amount of aspiration, the study was aborted and the patients were encouraged to expectorate the food material. All of the study procedures were recorded on AVI files (30 frames/second). After all of the patients finished the VFSS study, the video recordings were collected and each file was given a randomized number. Next, the files were copied to 10 DVDs with each DVD containing all of the video recordings in a different randomized order. The DVDs were sent to the interpreter for analysis.

Interpretation

All of the participating interpreters were physiatrists who had at least 5 years of experience in interpreting VFSS results. They agreed to participate after being informed of the nature of this study. All patient information, including age, sex, and underlying diseases, was withheld from the interpreters. The interpreters only observed the patients using the files on the DVD and described their findings using a standardized format (Appendix 1).

Statistical Methods

The intra-class correlation coefficient (ICC) model 2.1 of the VDS was calculated in order to test the inter-rater reliability based on the VDS scores provided by the interpreters. The ICC model was used because it can be used

not only for scale variables but also for ordinal variables. Ordinal variables equivalent to the weighted kappa ICC values over 0.80 was considered “very good”, and ICC values between 0.60-0.80 were considered “good”. The consistency of the other items was evaluated using Cohen’s kappa (κ).

RESULTS

One hundred patients (59 males and 41 females) with dysphagia were enrolled, including 64 stroke patients, 13 patients with traumatic brain injury, 12 patients with head and neck cancer, 6 patients with brain tumors, and 5 patients with other diseases. The average age of the enrolled patients was 64.4±14.8 years. All of the recruited patients underwent VFSS. Inter-rater reliability of the oral phase parameters are shown in Table 1. All of the oral phase parameters demonstrated low reliability ($\kappa < 0.4$). Among the oral phase parameters, lip closure showed the highest reliability ($\kappa = 0.325$), whereas premature bolus loss and oral apraxia demonstrated the lowest reliabilities ($\kappa = 0.060$ and $\kappa = 0.099$, respectively). Table 1 also presents data on pharyngeal phase reliability. Pharyngeal phase parameters demonstrated higher reliability than the oral phase parameters, but the κ value was below 0.4. Aspiration showed the highest reliability of all of the tested pa-

Table 1. Inter-rater Reliability of VDS

	K	SE	95% CI	
Lip closure	0.325	0.024	0.278	0.373
Bolus formation	0.153	0.023	0.108	0.197
Mastication	0.123	0.025	0.074	0.171
Apraxia	0.099	0.018	0.063	0.135
Tongue to palate contact	0.153	0.024	0.107	0.200
Premature bolus loss	0.060	0.016	0.028	0.092
Oral transit time	0.253	0.026	0.202	0.303
Triggering of pharyngeal swallowing	0.300	0.026	0.250	0.351
Vallecular residue	0.275	0.015	0.245	0.304
Laryngeal elevation	0.202	0.026	0.152	0.253
Pyramidal sinus residue	0.345	0.017	0.312	0.379
Coating on the pharyngeal wall	0.310	0.026	0.260	0.361
Pharyngeal transit time	0.165	0.026	0.115	0.216
Aspiration	0.393	0.019	0.356	0.431
	ICC*		95% CI	
Total score	0.556		0.463	0.648

SE: Standard error, CI: Confidential interval, ICC: Intra-class correlation coefficient

rameters ($\kappa=0.393$). Total score reliability, in terms of the ICC, was 0.556.

DISCUSSION

The past two decades have brought an enormous widening of our knowledge about dysphagia research and treatment.¹³ The most valuable and frequently used diagnostic tool for the evaluation of dysphagia is VFSS. While the VFSS protocol has been standardized for use in many research projects,¹¹ it also has a limited ability to predict dysphagia prognosis and provide the quantitative evaluation of dysphagia. Many physicians have tried to predict the long-term prognosis of dysphagia, and as a result, there are several studies on the long-term prognosis of dysphagia after a stroke. Delayed oral transit time, penetration, age over 70 years, poor Barthel index, and the presence of a frontal and insular cortex lesion have been suggested to indicate poor prognosis.^{7,14,15} However, if the risk factors alone cannot explain the quantitative probability of poor prognosis of dysphagia, then the VDS should be used to quantitatively investigate and predict the severity of dysphagia 6 months after the onset of a stroke.⁹

Overall, the VDS score demonstrated low to moderate reliability in our study (0.556 in terms of ICC). However, 14 individual sub parameters, particularly the oral phase parameters, showed low reliability. A previous study conducted by Stoeckli et al.¹⁰ reported low oral phase reliability ($\kappa=0.15-0.56$); the highest value was for lip closure ($\kappa=0.56$). Lip closure also demonstrated the highest reliability in our study ($\kappa=0.35$). Stoeckli et al.¹⁰ reported higher values than those of our study because lip closure was classified as a binary value (“yes” or “no”) in their study, without any intermediate values. Lip closure on VDS has 3 categorical values (“intact”, “inadequate”, and “none”); however, “inadequate” lip closure lacks an accurate definition and can be defined arbitrarily by the interpreter depending on which food material is used as the standard for evaluation. For example, if the lip closure of a patient was very good for a pureed diet but poor for the liquid diet, it might be classified as “inadequate” or “none” depending on which food material the interpreter chose to use as the standard.

Regarding the pharyngeal phase, the overall reliability was higher than the oral phase ($\kappa=0.165-0.393$ vs.

$\kappa=0.060-0.325$, respectively), similar to other studies that reported higher reliability for pharyngeal phase parameters than oral phase parameters.¹⁰ This is because many pharyngeal phase parameters have two categorical values (e.g., the triggering of pharyngeal swallowing, laryngeal elevation, the coating of the pharyngeal wall, pharyngeal transit time). Also, the pharyngeal phase parameters can be clearly seen by the VFSS. Penetration was defined as the passage of material into the larynx, but not through the vocal folds, and aspiration was defined as the passage of material through the vocal folds.¹⁶ These pharyngeal phase findings are relatively easier to differentiate than other oral phase findings.

The total VDS score demonstrated higher reliability than the individual parameters (0.556 in terms of ICC). This is due to the dilution effect of the scores of each parameter given by the interpreters.

The overall reliability is not particularly high in our study, and we believe this is because no clear definitions exist for intermediate values VDS, even though 9 of the 14 parameters have at least 3 categorical values. For example, “intact” mastication is given 0 points and “inadequate” mastication is given 4 points according to the VDS; however, depending on how each interpreter classifies the patient’s mastication function, a single patient can be given any point-- either 0 or 4. Therefore, the evaluation of patients showing some poor functioning of the parameters may lack consistency from interpreter to interpreter. Second, the guidelines specifying the type of food to be used as a standard for evaluation do not exist. In our study, various types of food material were tested on each patient. Depending on which type of material was used as the standard for evaluation, VFSS findings may be classified differently for each patient. For example, patients demonstrating good swallowing of solid foods but poor swallowing of liquid foods may be interpreted differently depending on whether solid or liquid foods was used for evaluation. For future studies, there should be guidelines regarding which food materials should be used as the standard for evaluating the findings related to each parameter.

This study has an obvious limitation. The interpretation was performed only via the observation of VFSS video recordings, as it was not logistically possible to have all 10 interpreters examine each patient. If the interpreters had been allowed to clinically examine their patients, this

would have improved the results of the interpretations by increasing accuracy. However, the object of this study was to evaluate inter-rater reliability of VDS based on VFSS findings. If the interpreter had predicted the findings from the clinical examination, this would have acted as a bias.

This is the first study to evaluate the inter-rater reliability of VDS. For future studies, a more precise and widely accepted study protocol will be needed. The development of such a protocol can be achieved by standardized education programs, such as interactive lecture movies or formal guidelines for interpreters. These education programs may contribute to achieving higher levels of accuracy in interpretation, and subsequently, to improving the abilities to predict the long-term prognosis of dysphagia.

CONCLUSION

VDS demonstrates a moderate rate of inter-rater reliability for evaluating the swallowing function. Some of the parameters demonstrated a lower rate of agreement, particularly the oral phase parameters. VDS has some limitations in predicting the long-term prognosis of dysphagia; hence, more accurate definitions of each parameter as well as a study protocol will be essential.

REFERENCES

1. Horner J, Massey EW. Silent aspiration following stroke. *Neurology* 1988; 38: 317-319
2. Linden P, Siebens AA. Dysphagia: predicting laryngeal penetration. *Arch Phys Med Rehabil* 1983; 64: 281-284
3. Eyal H. Evaluation and treatment of swallowing disorders (dysphagia)--the American model as an opportunity for improving patient care and cost containment in Israel's healthcare system. *Harefuah* 1997; 132: 775-778
4. Shapiro J. Evaluation and treatment of swallowing disorders. *Compr Ther* 2000; 26: 203-209
5. 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: 677-682
6. Paik NJ, Kim IS, Kim JH, Oh BM, Han TR. Clinical validity of the functional dysphagia scale based on videofluoroscopic swallowing study. *J Korean Acad Rehab Med* 2005; 29: 43-49
7. Mann G, Hankey GJ, Cameron D. Swallowing function after stroke: prognosis and prognostic factors at 6 months. *Stroke* 1999; 30: 744-748
8. Kidd D, Lawson J, Nesbitt R, MacMahon J. The natural history and clinical consequences of aspiration in acute stroke. *QJM* 1995; 88: 409-413
9. Han TR, Paik NJ, Park JW, Kwon BS. The prediction of persistent dysphagia beyond six months after stroke. *Dysphagia* 2008; 23: 59-64
10. Stoeckli SJ, Huisman TA, Seifert B, Martin-Harris BJ. Interrater reliability of videofluoroscopic swallow evaluation. *Dysphagia* 2003; 18: 53-57
11. Logemann JA. Manual for the videofluorographic study of swallowing, 2nd ed, Austin: Pro-Ed, 1993, 1-170
12. National Dysphagia Diet Task Force. National dysphagia diet: standardization for optimal care, 1st ed, Chicago: American Dietetic Association, 2002, 36
13. Logemann JA. Evaluation and treatment of swallowing disorders, 2nd ed, Austin: TX, 1998, 18
14. Broadley S, Croser D, Cottrell J, Creevy M, Teo E, Yiu D, Pathi R, Taylor J, Thompson PD. Predictors of prolonged dysphagia following acute stroke. *J Clin Neurosci* 2003; 10: 300-305
15. Mann G, Hankey GJ. Initial clinical and demographic predictors of swallowing impairment following acute stroke. *Dysphagia* 2001; 16: 208-215
16. Braddom RL. Physical medicine and rehabilitation, 4th ed, Philadelphia: Saunders, 2011, 586-587

Appendix 1. Videofluoroscopic Dysphagia Scale⁹

Parameter	Findings	Parameter	Findings	
Lip Closure	Intact	Triggering of pharyngeal swallow	Normal	0
	Inadequate		Delayed	4.5
	None			
Bolus formation	Intact	Vallecular residue	None	0
	Inadequate		<10%	2
	None		10-50%	4
				>50%
Mastication	Intact	Laryngeal elevation	Normal	0
	Inadequate		Impaired	9
	None			
Apraxia	None	Pyriform sinus residue	None	0
	Mild		<10%	4.5
	Moderate		10-50%	9
	Severe		>50%	13.5
Tongue to palate contact	Intact	Coating on the pharyngeal wall	No	0
	Inadequate		Yes	9
	None			
Premature bolus loss	None	Pharyngeal transit time	<1.0 s	0
	<10%		>1.0 s	6
	10-50%			
	>50%			
Oral transit time	<1.5 s	Aspiration	None	0
	>1.5 s		Penetration	6
			Aspiration	12