

Inter-rater Agreement for the Clinical Dysphagia Scale

Se Woong Chun, M.D., Seung Ah Lee, M.D., Il-Young Jung, M.D.,
Jaewon Beom, M.D., Tai Ryoan Han, M.D., Ph.D., Byung-Mo Oh, M.D., Ph.D.

Department of Rehabilitation Medicine, Seoul National University College of Medicine, Seoul 110-744, Korea

Objective To investigate the inter-rater agreement for the clinical dysphagia scale (CDS).

Method Sixty-seven subjects scheduled to participate in a video-fluoroscopic swallowing study (VFSS) were pre-examined by two raters independently within a 24-hour interval. Each item and the total score were compared between the raters. In addition, we investigated whether subtraction of items showing low agreement or modification of rating methods could enhance inter-rater agreement without significant compromise of validity.

Results Inter-rater agreement was excellent for the total score (intraclass correlation coefficient (ICC): 0.886). Four items (lip sealing, chewing and mastication, laryngeal elevation, and reflex coughing) did not show excellent agreement (ICC: 0.696, 0.377, 0.446, and κ : 0.723, respectively). However, subtraction of each item either compromised validity, or did not improve agreement. When redefining 'history of aspiration' and 'lesion location' items, the inter-rater agreement (ICC: 0.912, 0.888, respectively) and correlation with new videofluoroscopic dysphagia score (PCC: 0.576, 0.577, respectively) were enhanced. The CDS showed better agreement and validity in stroke patients compared to non-stroke patients (ICC: 0.917 vs 0.835, PCC: 0.663 vs 0.414).

Conclusion The clinical dysphagia scale is a reliable bedside swallowing test. We can improve inter-rater agreement and validity by refining the 'history of aspiration' and 'lesion location' item.

Key Words Deglutition disorder, Dysphagia, Reproducibility.

INTRODUCTION

Swallowing is a peristaltic wave of movements of the oropharyngeal parts, finely coordinated in a short period of time.^{1,2} Damage or pareses in the related parts, or imperfection of each movement may cause dysphagia.

The incidence of dysphagia is quite high after stroke or head and neck cancer surgery. Dysphagia gives rise to pneumonia, which is one of its typical complications, and leads to death of the patients, thus resulting in substantial socioeconomic loss.^{3,4}

This has encouraged efforts to develop a screening test to decide whether to promptly run a further test or apply treatment before the dysphagia causes pneumonia.^{5,6} In stroke unit settings, it has been recommended that screening tests be run as a preliminary study of a videofluoroscopic swallow study (VFSS) and VFSS should be done when the screening test turns out to be positive.^{7,8} However, most of the screening tests have low specificity and do not document severity.

Received December 14, 2010; Accepted February 8, 2011
Corresponding author: Byung-Mo Oh
Department of Rehabilitation Medicine, Seoul National University Hospital, 28, Yongon-dong, Chongno-gu, Seoul, 110-744, Korea
Tel: +82-2-2072-2619, Fax: +82-2-743-7473, E-mail: keepwiz@gmail.com
© 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 © 2011 by Korean Academy of Rehabilitation Medicine

The clinical dysphagia scale (CDS) is a dysphagia rating scale that can be used with ease at the bedside,⁹ which is a required condition of screening tests. It predicts the aspiration of patients with more precision, and can quantify the severity of dysphagia. It showed excellent sensitivity and specificity, and correlated well with VFSS findings.¹⁰ However, the ratings for some items, such as “history of aspiration” and “laryngeal elevation”, were somewhat ambiguous, raising concerns about its inter-rater agreement. This study aimed to investigate the inter-rater agreement of the CDS for the total score as well as each item score, and to explore possibilities of improving agreement by item modification if necessary.

MATERIALS AND METHODS

Subjects

Medical records of 133 patients (age \geq 18 years) with swallowing problems who underwent VFSS from June 29th to August 28th of 2009 in Seoul National University Hospital were reviewed for the study. In our hospital, to confirm whether a patient can safely undertake VFSS, it is our routine to visit the patient the previous day of the exam and check his/her condition as well as the CDS. The same procedure is repeated just before the exam by another physician. Among the reviewed records, 67 studies that had complete information on both CDS scores of two different raters and VFSS result data were analyzed in the study. The mean age \pm standard deviation of the subjects was 67.0 \pm 2.5 years and 32 were males. Thirty-seven were stroke patients whereas the others had dysphagia of different etiology (e.g: cardiac surgery, oropharyngeal cancer, inflammatory myopathies, and so on). The protocol for this retrospective study was approved by the Institutional Review Board of Seoul National University Hospital.

Raters

Every subject underwent VFSS and was pre-evaluated by two different raters. The first rater was a medical doctor who performed the rating within a day before VFSS, and had the ability to perform basic neurological examination. They were briefly (<1 hour) instructed on how to fill out the check-list for CDS after examining the patient. The doctor who performed the second CDS rating just before VFSS was a physiatrist with more than

2 years experience treating dysphagic patients, and was instructed in a similar way. Thus, inter-rater agreement could be tested based on the collected data.

The clinical dysphagia scale

The CDS consisted of 8 rating items (lesion location, tracheostomy, history of aspiration, lip sealing, chewing and mastication, tongue protrusion, laryngeal elevation, and reflex coughing).⁹ Lesion location indicated whether the location of ischemia/hemorrhage within the brain involved the brain stem, which was checked by medical record. If the etiology of dysphagia was not stroke, then it was not rated. Whether the patient had tracheostomy or not was identified by inspection. The rater asked the patient or caregiver whether the patient had experienced aspiration during the past week and rated the history of the aspiration item. If the patient had not tried oral feeding for the previous week due to nasogastric tube feeding or total parenteral nutrition, the item was not rated. Integrity of lip sealing, chewing and mastication, tongue protrusion, and laryngeal elevation was assessed by physical examination. These were rated according to three choices (intact, inadequate, and none). Reflex coughing was checked after allowing the patient to drink 3 ml of sterile water twice. In addition, we attempted various modifications of the CDS rating system in order to improve the inter-rater agreement without compromising the validity. The validity was checked by calculating the correlation of CDS and new Videofluoroscopic Dysphagia Scale (VDS).¹¹

The videofluoroscopic dysphagia scale (VDS)

The VDS was obtained based on VFSS results and VFSS was performed as follows: subjects were placed upright and given 2 and 5 mldiluted barium, pudding, rice gruel, yoplait, and boiled rice twice in a spoon. If the patient showed aspiration on the videofluoroscope or any clinical symptoms of aspiration, they progressed to the next test diet. The VDS was used as a standard when checking the validity of the CDS. The VDS was composed of 14 items that represented oral (lip closure, bolus formation, mastication, apraxia, premature bolus loss, and oral transit time) and pharyngeal function (pharyngeal triggering, vallecular and pyriform sinus residue, laryngeal elevation and epiglottic closure, pharyngeal coating, pharyngeal transit time, and

aspiration) observed in the VFSS. It was shown to have good correlation with the swallowing status of the patients.¹²

Statistical analysis

Intra-class correlation coefficient model 2,1 (ICC(2,1)) of the CDS was calculated, in order to test the inter-rater agreement based on the CDS scores by the two raters.¹³ ICC was also used to test the consistency of ordinal items with three choices. The ICC can be used in both scale and ordinal variables. Also, the meaning of ICC in ordinal variable is equivalent to that of weighted kappa.¹⁴ An ICC higher than 0.80 was considered 'excellent'.¹³ The consistency of other items was evaluated by Cohen's kappa (κ) because they were categorical binomial variables. A κ higher than 0.60 was defined as 'good' agreement.¹⁵ Correlation between CDS and VDS was assessed by Pearson's correlation coefficient (PCC). When modifying the CDS, a PCC value under 0.489 was considered as compromising the validity of the test based on the previous study results.¹⁰ In all the three statistical methods, zero meant nil correlation between two variables. A p-value <0.05 was considered statistically significant.

RESULTS

The CDS showed excellent inter-rater agreement (ICC (95% confidence interval): 0.886 (0.814-0.930), $p < 0.001$). Although five items (lesion location, tracheostomy, history of aspiration, reflex coughing, and tongue protrusion) showed good agreement (κ : 0.735, 1.000, 0.802, 0.723, and ICC: 0.837, respectively,) the other three items (lip sealing, chewing and mastication, and laryngeal elevation) did not (ICC: 0.696, 0.377, and 0.446, respectively) (Table 1). The CDS total score also showed significant correlation with VDS (PCC: 0.560, $R^2 = 0.3136$, $p < 0.001$).

We attempted to improve inter-rater agreement by excluding items showing low agreement. The ICC of the total sum of the remaining seven items after subtracting lesion location, lip sealing, chewing and mastication, laryngeal elevation, and reflex coughing (the items that did not show excellent agreement, $ICC > .80$, $\kappa > 0.80$) was 0.883, 0.886, 0.881, 0.895, and 0.956, respectively. Subtracting the last item improved the agreement considerably, ICC: 0.956 (0.928-0.973) vs. 0.886 (0.814-0.930), only with compromise of the correlation with VDS (PCC: 0.266) (Table 2).

Table 1. Inter-rater Agreement of the Clinical Dysphagia Scale and Correlation with the Videofluoroscopic Dysphagia Scale

Item	Total (n=67) ICC (95% CI)	κ (% Agree)	Stroke (n=37) ICC (95% CI)	κ (% Agree)	Non-stroke (n=30) ICC (95% CI)	κ (% Agree)
Total score	.886 [†] (.814-.930)		.917 [†] (.839-.957)		.835 [†] (.656-.921)	
Lip sealing	.696 [†] (.506-.813)		.747 [†] (.510-.870)		.606 [†] (.162-.813)	
Chew/mastication	.377* (-.019-.619)		.280 (-.425-.633)		.446 (-.189-.739)	
Tongue protrusion	.837 [†] (.734-.900)		.917 [†] (.839-.957)		.763 [†] (.501-.887)	
Larynx elevation	.446 [†] (.116-.655)		.492* (.036-.735)		.404 (-.189-.709)	
Lesion location		.735 [†] (97.0)		.841 [†] (97.1)		NA
Tracheostomy		1.000 [†] (100)		1.000 [†] (100)		1.000 [†] (100)
Aspiration		.802 [†] (91.0)		.709 [†] (89.2)		1.000 [†] (100)
Reflex cough		.723 [†] (86.6)		.753 [†] (89.2)		.670 [†] (83.3)

ICC: Intraclass correlation coefficient, 95% CI: 95% Confidence interval, % Agree: Percentage of agreement, NA: Not assessed, *p-value<.05, [†]p-value<.01

Table 2. Inter-rater Agreement of the Clinical Dysphagia Scale after Subtraction of One Item and Correlation with the Videofluoroscopic Dysphagia Scale

Items to be subtracted	Inter-rater agreement		Correlation with VDS
	ICC	95% CI	PCC
Location	.883	.810-.928	.563
Tracheostomy	.846	.750-.905	.582
Aspiration	.890	.822-.932	.539
Lip sealing	.886	.815-.930	.564
Chewing and mastication	.881	.808-.927	.549
Tongue protrusion	.884	.811-.929	.553
Laryngeal elevation	.895	.829-.935	.541
Reflex coughing	.956	.928-.973	.266

ICC: Intraclass correlation coefficient, 95% CI: 95% Confidence interval, PCC: Pearson's correlation coefficient, VDS: Videofluoroscopic dysphagia scale

Table 3. Inter-rater Agreement and Validity after Modification

	Inter-rater agreement		Correlation with VDS
	ICC	95% CI	PCC
Original	.886	.814-.930	.560
Modified item			
History of aspiration	.912	.857-.946	.576
Lesion location	.888	.818-.931	.577
Both combined	.913	.858-.946	.589

ICC: Intraclass correlation coefficient, 95% CI: 95% Confidence interval, PCC: Pearson's correlation coefficient

We modified the rating means of the 'history of aspiration' item. If the patient had been on nasogastric tube feeding or total parenteral nutrition and not tried oral feeding for a week, then the rating was changed from equivalent to 'yes.' Thus, the altered CDS showed better inter-rater agreement and validity (ICC: 0.912 (0.857-0.946), $p < 0.001$; PCC: 0.576, $p < 0.001$). We also attempted to refine the 'lesion location' item. If the patient did not have stroke, we rated it as equivalent to stroke involving the brain stem. This modification showed similar inter-rater agreement and improved validity (ICC: 0.888 (0.818-0.931), $p < 0.001$; PCC: 0.577, $p < 0.001$). Combining the modifications showed additional increment in validity (ICC: 0.913 (0.858-0.946), $p < 0.001$; PCC: 0.589, $p < 0.001$) (Table 3).

Patients were grouped according to whether the etiology

of dysphagia was stroke or not. The CDS showed better agreement and validity in stroke patients compared to non-stroke patients (ICC: 0.917 (0.839-0.957), $p < 0.001$ vs 0.835 (0.656-0.921), $p < 0.001$; PCC: 0.663, $p < 0.001$ vs 0.414, $p < 0.001$) (Table 1).

DISCUSSION

A valid screening test for dysphagia is fundamental for the discrimination of patients with swallowing problems in the acute phase of disease. Thus, we can promptly decide on how to provide nutrition without increasing the risk of aspiration pneumonia or causing unnecessary discomfort of nasogastric tube feeding. Therefore, many screening tests for dysphagia have carried out for a long time.¹⁶ The 90 ml (3 oz) water swallow test was introduced in 1992¹⁷ and validated for use in stroke patients.¹⁸ In 1998, speech language pathologists investigated a bedside assessment tool and assess whether it could predict aspiration.¹⁹ They assessed head posture, trunk control, drowsiness, communication, lip closure, tongue movement, gag reflex, coughing and drinking. This test showed 47% sensitivity and 86% specificity with moderate or less than moderate inter-rater agreement. Hinds et al. disclosed the results of the water swallow test, which allowed patients to swallow 100-150 ml of water, beginning with a small amount.²⁰ It showed 97% sensitivity and 69% specificity. However, the outcome measure was not a direct confirmation of aspiration such as aspiration ascertained in VFSS. Due to the limitations of these previous studies, new dysphagia screening tests were introduced. Recently, the Gugging swallowing screening⁵ and the Toronto Bedside Swallowing Screening Test (TOR-BSST)⁶ were reported to have high sensitivity. However, the recently introduced tests tend to show low specificity.

The CDS was developed from a group of 59 stroke patients with an average age of 63 years. It was developed to predict aspiration ascertained by VFSS. The eight items were selected among various clinical findings using a polychotomous linear logistic regression model using aspiration as a criterion factor and various clinical findings as predictor factors. Eight clinical findings with statistical significance were selected as CDS items. Each item was given weight based on the odds ratio so that the total score would be 100 points, higher score indicating higher probability of aspiration.⁹ With a cut-value of 40

points, it showed excellent sensitivity and specificity.¹⁰

The present study demonstrated that the CDS showed excellent inter-rater agreement. Although inter-rater agreement of some individual items was low, a sum CDS showed an excellent level of agreement. All items that failed to show good inter-rater agreement (lip sealing, chewing and mastication, and laryngeal elevation) were those that were rated by physical examination. This is inevitable when two raters of different experience examine a patient. Moreover, the two exams of different raters lacked temporal synchronicity, although the time lapse was just about 24 hours. More thorough education and training rather than brief one-hour instruction may improve inter-rater agreement. However, this will make the CDS less applicable in a clinical environment, which takes away a major strength of the CDS rating system. Considering that reflex coughing is an obvious sign, its low ICC value may seem peculiar. The 'positive' reflex coughing was defined as coughing or wet voice during three trials of drinking 3 ml sterile water. The wet voice criterion may have affected the consistency. Many patients who were referred to undertake VFSS had poor lung condition. Therefore, wet voice due to excessive sputum and poor expectoration might have confounded the examination. Examination after complete throat clearing or removal of wet voice would improve inter-rater agreement.

Subtraction of each item was done to improve inter-rater agreement. It failed each time except for reflex coughing. However, validity was compromised when it was subtracted. Reflex cough is a direct sign of aspiration. Therefore, it is obvious that this item closely correlates with VDS, which contains an aspiration item.

Concerns have been raised over the vagueness of the rating method on some items. There has been no clear guideline on the items that cannot be rated. For example, patients who had been under the order 'nil per os' in the preceding weeks, cannot be rated adequately for the 'history of aspiration' item. Therefore, we modified the rating method of two items under the judgment that the established method lacks logical foundation. Nasogastric tube feeding or total parenteral nutrition were indicated when a patient was in various acute medical conditions or had severe difficulty in swallowing. In the acute care setting, the clinician looks for any sign of aspiration and if the patient shows any, oral nutrition is usually forbidden.

Therefore, we rated having nasogastric tube feeding/total parenteral nutrition and no oral feeding as equivalent to having experience with aspiration. This attempt improved the validity as well as the agreement.

We also clarified a guideline on the 'lesion location' item. Although over 50% of stroke patients show dysphagia in their acute phase, a small percentage of survivors suffer from chronic swallowing problems.²¹ Dysphagia caused by stroke involving the unilateral brain is typically transient with the exception of stroke that involves the brain stem.²² On the contrary, non-stroke origin dysphagia is due to permanent anatomic change after surgery, chemoradiation, deteriorated general condition, neurodegenerative disease, or muscle disease. Therefore, dysphagia is sustained or progresses in many cases. Hence, we modified the rating method of the 'lesion location' item. Having dysphagia with etiology other than stroke was considered equivalent to stroke involving the brain stem. The same tendency of improvement was observed in both validity and inter-rater agreement.

The CDS was originally designed for stroke patients. Therefore, it is quite obvious that CDS shows better correlation with VFSS findings in stroke patients compared to non-stroke patients. This fact did not change even after removal of the 'lesion location' item, which indicated the location of stroke. This would be due to the different mechanism of dysphagia. In stroke patients, the reason for aspiration is impaired sensory input, paresis, and incoordination of swallowing muscles. On the contrary, in non-stroke patients the pathomechanism of dysphagia can be very different. For example, a patient may have partial laryngectomy and develop dysphagia despite good oral function and laryngeal elevation. There is no item that can assess this in the current CDS rating system.

CONCLUSION

The CDS was revealed to have excellent inter-rater agreement. Although some items do not show as much agreement, such as the total score, the CDS is a reliable rating system. To sum up, CDS is an adequate screening tool that can be easily learned and applied by physicians even without rich experience in dysphagia treatment for reliable detection of dysphagia, and for the selection of

patients who should undertake VFSS. Modification of some of the items improved the agreement and validity. Accordingly, we suggest a revised version of CDS with short instruction (Appendix 1).

REFERENCES

1. Cook IJ, Dodds WJ, Dantas RO, Kern MK, Massey BT, Shaker R, Hogan WJ. Timing of videofluoroscopic, manometric events, and bolus transit during the oral and pharyngeal phases of swallowing. *Dysphagia* 1989; 4: 8-15
2. Perlman AL, Palmer PM, McCulloch TM, Vandaele DJ. Electromyographic activity from human laryngeal, pharyngeal, and submental muscles during swallowing. *J Appl Physiol* 1999; 86: 1663-1669
3. Koivula I, Sten M, Makela PH. Risk factors for pneumonia in the elderly. *Am J Med* 1994; 96: 313-320
4. Niederman MS, McCombs JS, Unger AN, Kumar A, Popovian R. The cost of treating community-acquired pneumonia. *Clin Ther* 1998; 20: 820-837
5. Trapl M, Enderle P, Nowotny M, Teuschl Y, Matz K, Dachenhausen A, Brainin M. Dysphagia bedside screening for acute-stroke patients: the Gugging Swallowing Screen. *Stroke* 2007; 38: 2948-2952
6. Martino R, Silver F, Teasell R, Bayley M, Nicholson G, Streiner DL, Diamant NE. The Toronto Bedside Swallowing Screening Test (TOR-BSST): development and validation of a dysphagia screening tool for patients with stroke. *Stroke* 2009; 40: 555-561
7. Bates B, Choi JY, Duncan PW, Glasberg JJ, Graham GD, Katz RC, Lamberty K, Reker D, Zorowitz R. Veterans Affairs/Department of Defense Clinical Practice Guideline for the Management of Adult Stroke Rehabilitation Care: executive summary. *Stroke* 2005; 36: 2049-2056
8. Duncan PW, Zorowitz R, Bates B, Choi JY, Glasberg JJ, Graham GD, Katz RC, Lamberty K, Reker D. Management of Adult Stroke Rehabilitation Care: a clinical practice guideline. *Stroke* 2005; 36: e100-143
9. Han TR, Paik NJ, Park JW. The clinical functional scale for dysphagia in stroke patients. *Korean J Stroke* 2001; 3: 153-157
10. Jung SH, Lee KJ, Hong JB, Han TR. Validation of clinical dysphagia scale: based on videofluoroscopic swallowing study. *J Korean Acad Rehab Med* 2005; 29: 343-350
11. Han TR, Paik NJ, Park JW, Kwon BS. The prediction of persistent dysphagia beyond six months after stroke. *Dysphagia* 2008; 23: 59-64
12. 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
13. Shrout PE, Fleiss JL. Intraclass correlations: uses in assessing rater reliability. *Psychol Bull* 1979; 86: 420-428
14. Fleiss JL, Cohen J. The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability. *Educ Psychol Meas* 1973; 33: 613-619
15. Fleiss JL, Nee JC, Landis JR. Large sample variance of kappa in the case of different sets of raters. *Psychol Bull* 1979; 86: 974-977
16. Splaingard ML, Hutchins B, Sulton LD, Chaudhuri G. Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment. *Arch Phys Med Rehabil* 1988; 69: 637-640
17. DePippo KL, Holas MA, Reding MJ. Validation of the 3-oz water swallow test for aspiration following stroke. *Arch Neurol* 1992; 49: 1259-1261
18. DePippo KL, Holas MA, Reding MJ. The Burke dysphagia screening test: validation of its use in patients with stroke. *Arch Phys Med Rehabil* 1994; 75: 1284-1286
19. Smithard DG, O'Neill PA, Park C, England R, Renwick DS, Wyatt R, Morris J, Martin DF. Can bedside assessment reliably exclude aspiration following acute stroke? *Age Ageing* 1998; 27: 99-106
20. Hinds NP, Wiles CM. Assessment of swallowing and referral to speech and language therapists in acute stroke. *QJM* 1998; 91: 829-835
21. Mann G, Hankey GJ, Cameron D. Swallowing function after stroke: prognosis and prognostic factors at 6 months. *Stroke* 1999; 30: 744-748
22. Kruger E, Teasell R, Salter K, Foley N, Hellings C. The rehabilitation of patients recovering from brainstem strokes: case studies and clinical considerations. *Top Stroke Rehabil* 2007; 14: 56-64

Appendix 1. Clinical Dysphagia Scale (revised)

Location*	Non-stem lesion	0
	Stem lesion	5
	Non-stroke etiology	5
T-cannula	No	0
	Yes	25
Aspiration [†]	No	0
	Yes	10
	Have not tried oral feeding	10
Lip sealing	Intact	0
	Inadequate	2
	None	4
Chewing and mastication	Intact	0
	Inadequate	4
	None	8
Tongue protrusion	Intact	0
	Inadequate	4
	None	8
Laryngeal elevation	Intact	0
	Inadequate	5
	None	10
Reflex coughing	No	0
	Yes	30
Total		

*If the patients' stroke lesion involves the brain stem, rate 5 points. If the etiology of dysphagia is other than stroke, rate 5 points. [†]If the patient had any aspiration symptoms or had not tried oral feeding in the past week, rate 10 points