Sensitivity And Specificity Of Clinical/Bedside Examination Signs For Detecting Aspiration In Adults Subsequent To Stroke

Authors
G.H. McCullough R.T. Wertz, J.C. Rosenbek

Abstract
While detecting the presence of aspiration is only one aspect of a noninstrumented examination of swallowing function, it is an important component due to its potential health status consequences. The purpose of this investigation was to investigate the sensitivity and specificity of clinical/bedside examination signs for predicting aspiration on videofluoroscopic examination of swallowing. Sixty consecutive, acute stroke patients were investigated with clinical/bedside and videofluoroscopic exams. Clinical signs consisted of history, oral motor/speech praxis, voice, and trial swallow ratings. Results confirm that more work needs to be done if data collected from noninstrumented examinations are to be strongly predictive of aspiration on VFSE. However, comparisons of the current results with previous investigations provides a promising framework for future research. Educational objectives: (1) To understand the use of sensitivity and specificity in detecting disease; (2) To understand the current evidence regarding clinical signs of aspiration.

G.H. McCullough, R.T. Wertz, J.C. Rosenbek (2001) "Sensitivity And Specificity Of Clinical/Bedside Examination Signs For Detecting Aspiration In Adults Subsequent To Stroke" *Journal of Communication Disorders* 34 pp. 55- 72 Version Of Record Available At www.sciencedirect.com

1. Introduction

Because of its possible health status consequences, aspiration is one of the most critical signs of oropharyngeal dysphagia. Therefore, identifying its presence in the overall profile of oropharyngeal dysphagia is an important clinical obligation. Videofluoroscopy is traditionally recognized as the evaluative tool of choice for this task, although a databased case for the equality of endoscopy is gathering momentum. The noninstrumented, minimally invasive clinical/bedside examination (C/B E) to elicit the signs of aspiration or signs and history posited to predict aspiration on instrumented examination is often criticized for its lack of sensitivity and specificity (Arvedson, Rogers, Buck, Smart, & Msall, 1994; Garon, Engle, & Ormiston, 1996; Horner, Massey, Riski, Lathrop, & Chase, 1988; Leder, Sasaki, & Burrell, 1998; Linden & Siebens, 1983; Splaingard, Hutchins, Sulton, & Chaudhuri, 1988). Sensitivity of a clinical sign for detecting aspiration can be defined as the proportion of patients who aspirate who are also positive for the clinical sign. Specificity of a clinical sign can be defined as the proportion of patients who do not aspirate who are also negative for the clinical sign (Friedland et al., 1998). Low sensitivity and specificity result, in part, from the fact that aspiration, in and of itself, can occur without detection, even by the most skilled examiner. Such aspiration has come to be called silent aspiration. Silent aspiration is estimated to occur in approximately 40% of all neurologic patients who aspirate (Logemann, 1998; Splaingard et al., 1988). Other recent research suggests that in addition to medical diagnosis, age and gender are also significantly correlated with the presence of silent aspiration (Smith, Logemann, Colangelo, Rademaker, & Pauloski, 1999). For a variety of reasons, including inadequate sensitivity and specificity data, the surging popularity of the C/B E, radiation exposure, and the difficulty of getting VFSE in some settings, the search for clinical signs to detect and/or predict aspiration continues, especially in stroke patients.

An abnormal, volitional cough (Daniels et al., 1998; Gordon, Hewer, & Wade, 1987; Horner, Brazer, & Massey, 1993; Horner, Massey, & Brazer, 1990) and the absence of a pharyngeal gag reflex (Daniels et al., 1998; Gordon et al., 1987; Horner et al., 1988; Linden & Siebens, 1983; Logemann, Veis, & Colangelo, 1999) have been identified as signs of aspiration in stroke patients by some researchers. Others have found no significant relationship between an abnormal, volitional cough or the lack of a pharyngeal gag reflex and aspiration (DeJong, 1967; Leder, 1997; Linden, Kuhlemeier, & Patterson, 1993).

Other signs, however, garner more consistent support from the data. Signs of laryngeal dysfunction, such as an overall rating of the presence or absence of dysphonia, have been identified in several studies (Daniels et al., 1998; Horner et al., 1988, 1990, 1993; Linden et al., 1993). Additional signs linked to aspiration in adults with neurologic etiologies are: the presence of dysarthria (Daniels et al., 1998; Hartelius & Svensson, 1994); depressed mental status (Chokshi, Asper, &

Khandheria, 1986; Feinberg, Ekberg, Segall, & Tully, 1992); cough after the swallow (Daniels et al., 1998; Logemann et al., 1999); voice change after the swallow (Daniels et al., 1998; Logemann et al., 1999); reduced laryngeal elevation (Logemann et al., 1999); multiple swallows per bolus (Logemann et al., 1999); difficulty managing secretions (Linden et al., 1993); and choking during the ""3 oz Swallow Test" (DePippo, Holas, & Reding, 1992). A history of pneumonia may also predict aspiration on VFSE (Cogen & Weinryb, 1989; Logemann et al., 1999).

For a variety of reasons, despite the relative unanimity of results, it is difficult to know what to conclude from these data. One problem is that some researchers did not compute sensitivity and specificity. Another is that despite the preponderance of stroke patients in these samples, enrollment criteria were dissimilar. It can be hypothesized that sensitivity and specificity will differ depending on whether patients are enrolled consecutively or selectively and whether they are referred by another healthcare professional or after failing a screening completed by a dysphagia clinician.

Additional complications are created by a series of assumptions that appear to underlie selection and use of procedures for the C/B E. Inclusion in the C/B E of many tasks also traditionally part of standardized perceptual speech testing, such as rapid syllable repetition, imply that researchers considered these tasks to be as important in dysphagia as they are in understanding dysarthria. As a result the C/B E is longer and, perhaps, less focused than it need be. More critical is the dearth of inter- and intrajudge reliability data for both instrumented and noninstrumented examinations. The assumption appears to be that judges can reliably evaluate responses to VFSE and C/B E procedures, although evidence of reliability problems for both the VFSE (Ekberg et al., 1988; Kuhlemeier, Yates, & Palmer, 1998; McCullough et al., in press; Scott, Perry, & Bench, 1998; Wilcox, Liss, & Siegel, 1996) that are improved with training (Scott et al., 1998) and the C/B E are accumulating (McCullough et al., 2000). Moreover, limited control over the kinds and viscosities of boluses swallowed at the bedside and during VFSE seems to betray an assumption that such control is not critical.

Given the importance of establishing the C/B E's place in dysphagia's armamentarium, programmatic efforts to test all such assumptions are necessary, and the present study is but one component of such an effort. The first component was an investigation of clinician preferences for and uses of an array of VFSE bolus sizes and viscosities and C/B E procedures (McCullough, Wertz, Rosenbek, & Dinneen, 1999). This was followed by two studies of inter- and intrajudge reliability, one for the C/B E (McCullough et al., 2000) and one for the VFSE (McCullough et al., in press). The present study investigated the sensitivity and specificity of signs and historical information elicited by commonly employed C/B E procedures and medical history questions (McCullough et al., 1999) for predicting the presence or absence of aspiration on VFSE in patients who had suffered an acute stroke.

2. Method

2.1. Patients

Sixty patients who suffered a recent thromboembolic stroke were entered in the study. Sample size was determined by power analysis to provide a power of .80 with an **a** at .05. Patients were recruited from the Veteran's Administration Medical Centers in Nashville and Murfreesboro, Tennessee and Vanderbilt University Medical Center in Nashville, Tennessee. Inclusion criteria for participation in the study were: occurrence of a stroke within 6 weeks of the time of examination, and mental competence to provide informed consent. Exclusion criteria were: presence of a structural anomaly that could interfere with swallowing, presence of or recent history of a tracheostomy, a reported history of dysphagia prior to the stroke, and/or physician judgement that the patient was too medically unstable to participate in the study.

Descriptive data on the 60 patients are shown in Table 1. Fifty-five of the sixty patients were male. The mean age of the patients was 67.8 years with a range of 40-96 years and a standard deviation of 9.94 years. The mean number of days post-onset from the most recent CVA was 5.98 days with a range of 1-42 days and a standard deviation of 7.38 days.

Brain imaging data indicated that 44 patients had suffered a single stroke, and 16 had suffered two or more strokes. Twenty-nine patients had sustained a cortical stroke or strokes. Eleven of these involved the right hemisphere, seventeen involved the left hemisphere, and one was bilateral. Fourteen patients had sustained a subcortical stroke or strokes. Nine of these were on the right side, four were on the left, and one was bilateral. Six patients had sustained a brainstem stroke or strokes. Two of these were on the right side, two were on the left, and two were bilateral. Three patients had sustained a cerebellar stroke, three had a mixed localization, and the lesion in three others could not be localized.

2.2. Design

A C/B E of swallowing and a VFSE were administered to each patient. Protocols for both were developed from a survey of clinicians who regularly evaluate adults with neurologic etiologies for dysphagia (McCullough et al., 1999).

Table 1
Descriptive data on study patients

Variable	*N*	Mean	Range	S.D.
Age in years	60	67.8	40-96	9.94
Days post-onset of CVA	60	5.98	1-42	7.38
Gender				
Male	55			
Female	5			

2.3. ClinicalIbedside examination (CIB E)

After informed consent was obtained, the primary study clinician performed a C/B E on each patient. There were four sections to the C/B E: history, oral motor (which included speech and praxis), voice, and trial swallows. Oral motor, voice, and swallowing performance was scored with a binary system, either +/− or normal-abnormal. Ratings of clinical signs in this investigation were made according to reports regarding clinicians' preferences and practices for evaluating dysphagia in adults (McCullough et al., 1999). There was no training to criterion for those ratings because clinicians typically are not trained to criterion before examining their patients (McCullough et al., 1999). Historical information was obtained from medical charts, physicians, nurses, patients, or families, depending on the sign. The presence or absence of surgery may indicate any surgery other than head/neck surgery, as that was one of the exclusion criterion for participation in the study. Additionally, poor nutrition was derived from the medical chart based on the evaluation of a registered dietician, not direct examination by the clinician. Decreased mental status was defined by historical information, neurologic examination, and clinical impression. Oral motor measures primarily related to the strength and range of motion of the oral structures involved in oral preparation and bolus propulsion. Also tested were palatal and pharyngeal gag reflexes, oral praxis, and ratings of dysarthria and speech intelligibility from a speech sample elicited by picture description. All oral motor signs were based on clinical impression, not information from the medical record. Voice measures were elicited by having the patient sustain an /a/ and by having him or her read The Grandfather Passage or discuss work history. For the trial swallows portion of the C/B E, two swallows of each consistency — thin liquid, thick liquid, puree, and solid — were administered in 5 cc boluses (1/4 cookie for solid). Thin and thick liquids were administered from a pill cup; puree and solids were administered from a spoon. Delayed swallow, total swallow duration, laryngeal elevation, and swallows per bolus were rated using the four-finger method (Logemann, 1998).

2.4. Videofluoroscopic examination (VFSE)

Within 24 h, after the completion of the C/B E, the primary study clinician conducted a VFSE. Patients were seated upright in a wheelchair or stretcher chair for the duration of the study. At the Nashville VA Medical Center, the studies were conducted with a mobile C-arm X-ray (model 9400) system (OEC-Diagnostics, Salt Lake City, Utah), run by a radiology technologist. Each study was recorded with a Panasonic Super VHS AG-1960 Pro Line Multiplex videocassette recorder with an attached digital videotimer (TEL Video Products — Model VC436). At the Murfreesboro VA Medical Center,

the study was conducted with a Phillips Super 80CP fluoro unit and recorded with a Panasonic A66300 VCR. At Vanderbilt University Medical Center, the study was conducted on a Siemens fluoro unit (Model #8842437G5275) with a 40 in. fixed tower and recorded with a Panasonic AG6300 MD videocassette recorder.

Examination began with two swallows of thin liquid — each 5 cc with a viscosity 14 cP (centipoise). This was a mixture of water and barium (E-Z-HD Barium Sulfate Powder for Suspension). These swallows were followed by two 10-cc swallows of thin liquid with the same viscosity. All thin liquid swallows were viewed laterally. The patient then swallowed two 5-cc boluses of thick liquid, each with a viscosity of about 187 cP. This was a mixture of thickened juice and barium (E-Z-HD Barium Sulfate Powder for Suspension). Each was viewed laterally. These swallows were followed by two 10-cc swallows of thick liquid with the same viscosity — each viewed laterally. Next, the patient was turned for a posterior-anterior view for two more 5-cc swallows of thick liquid to allow a judgement of bolus flow through the right versus left pyriform sinuses. The patient was then returned to the lateral view. In this view the patient swallowed two 5-cc boluses of applesauce (mixed with E-Z-HD Barium Sulfate Powder for Suspension) from a spoon followed by two solids (1/4 Lorna Doone cookie coated with Barium Sulfate Esophageal Cream).

At least 1 week after the completion of the videofluoroscopic examination and when five study tapes had been collected, the primary study clinician viewed the videotaped videofluoroscopy studies, blinded to information regarding the patient and his or her C/B E, and made his ratings. Only the rating of aspiration as present or absent is discussed in this article.

2.5. Reliability

Inter- and intrajudge reliability for both the C/B E signs and the VFSE measures were obtained on 25% ($N = 15$) of the patients studied. For the C/B E, interjudge reliability measures were obtained by having a second clinician make ratings and judgements along with the primary study clinician. The clinicians did not discuss the judgements that they made during or after the examination. Intrajudge reliability was obtained by having the primary study clinician administer a second C/B E on the patient the day after the initial evaluation. For patients on whom reliability data were derived, the VFSE was not conducted until both C/B Es had been completed. For intrajudge reliability of VFSE measures, the primary study clinician reviewed each tape a second time — at least 1 week after the original viewing, and recorded all measurements on a new data sheet. Interjudge reliability was assessed by having a second clinician view the same videofluoroscopy studies and complete the data sheets separately. The second clinician had over 200 h of experience with C/B E and VFSE examinations of swallowing.

2.6. Statistical analysis

A signal detection analysis program (Chial, 1997) was used to determine sensitivity, specificity, A prime (A^t), and chi-square (χ^2). Presence or absence of relevant history items and signs elicited during the C/B E were compared with the presence or absence of aspiration as determined by the VFSE. To do so, a series of two-by-two contingency tables was utilized. If aspiration was present on VFSE when a history item or sign was present on the C/B E, a ""true positive" rating resulted. If aspiration was absent when a history item or sign was absent, a ""true negative" rating resulted. If aspiration was not present on VFSE but the C/B E sign was present, a ""false positive" rating resulted. If aspiration was present on the VFSE but the C/B E sign was absent, a ""false negative" rating resulted. Sensitivity for each C/B E item was computed by dividing the number of patients with a true positive by the number of patients with a true positive plus the number with a false negative. Specificity for each C/B E item was computed by dividing the number of patients with a true negative by the number of patients with false positive plus the number with a true negative. There were 55 2×2 tables in all. The closer the sensitivity value for each C/B E item is to 1.000, the better its sensitivity for detecting the presence of aspiration, as confirmed by VFSE. The same is true for specificity. Sensitivity and specificity values of .600 or better were arbitrarily chosen to represent a minimal level of acceptable sensitivity and specificity (Chial, 1998).

A^t is a ""non-parametric index of test performance as a detector of disease, independent of test criterion" (Chial, 1997). The values for A^t can range from .500, which indicates the test is not a good detector of disease, to .999, which means the test is a phenomenal detector of disease. An A^t of .700 was arbitrarily chosen to represent a minimal level of acceptability (Chial, 1998). Chi-square values that were significant at $P < .05$ were considered to be acceptable. In sum, to meet a level of minimal acceptability, a clinical sign should have a sensitivity and specificity of .600 or greater, an A^t of .700 or greater, and a chi-square that is significant at $P < .05$ for detecting the presence or absence of aspiration on VFSE. All values for all C/B E items are reported.

For analysis of inter- and intrajudge reliability of the binary ratings for C/B E items and the VFSE measure of aspiration, Cohen's kappa was employed.

3. Results

3.1. Videofluoroscopic (VFSE) results

Aspiration observed in the VFSE was used as the gold standard to determine the sensitivity and specificity for each clinical sign. Twenty-two patients (37%) were judged to aspirate. Table 2 shows the relationship between aspiration in the VFSE and site of lesion. Nine of twenty-nine patients (31%) with a cortical

Table 2
Lesion localization and the number of patients who did not aspirate

Localization	*N*	Aspiration	No aspiration
Cortical			
Right	11	4	7
Left	17	5	12
Bilateral	1		1
Subcortical			
Right	9	3	6
Left	4	1	3
Bilateral	1	1	
Brainstem			
Right	2	1	1
Left	2	1	1
Bilateral	2	1	1
Cerebellar	3	1	2
Mixed	5	1	4
Questionable	3	3	

lesion, five of fourteen patients (36%) with a subcortical lesion, three of six patients (50%) with a brainstem lesion, one of three patients (33%) with a cerebellar lesion, one of five patients (10%) with mixed localization, and three of three patients (100%) with undetermined localization aspirated during the VFSE.

3.2. Reliability

Intra- and interjudge reliability for the C/B E signs are located in Tables 3–6 under the columns ""INTRA" and ""INTER," respectively. A ""+" next to a measure indicates that Cohen's kappa was significant at $P < .05$. A ""–" next to a measure indicates that Cohen's kappa was not significant at $P < .05$. Eight of the fifteen history measures were rated with significant interjudge reliability (see Table 3). Intrajudge reliability for history measures was not analyzed because the primary study clinician would have been able to extract information from the same sections of the medical chart, thus providing misleading reliability. Interjudge reliability was not assessed for ""nutritional status" because the information was obtained from a dietician. All oral motor measures were rated with significant inter- and intrajudge reliability except for the presence or absence of a palatal gag, which lacked both inter and intrajudge reliability and lip retraction, which lacked interjudge reliability (see Table 4). Eight of the twelve voice measures were rated with significant inter- and intrajudge reliability, including an overall rating of the presence or absence of dysphonia (see Table 5). Six of the ten trial swallow measures were rated with significant inter- and intrajudge reliability (see Table 6).

Inter- and intrajudge reliability for rating the presence or absence of aspiration from VFSEs was significant ($P < .05$) for all bolus sizes and consistencies.

Table 3
Reliability, sensitivity, specificity, A^t, and chi-square for history components of a clinical/bedside exam for detecting aspiration

History	INTRA	INTER	SENS	SPEC	A^t	x^2
1. Patient report	N/A	+	.381	.789	.619	0.861
2. Family report	N/A	+	.136	.789	.341	0.512
3. Nurse report	N/A	−	.318	.789	.619	0.861
4. Pneumonia	N/A	+	.318	.921	.753	5.742 *
5. GI disorder	N/A	−	.227	.684	.370	0.538
6. Surgery	N/A	−	.364	.579	.436	0.191
7. Related disease	N/A	−	.905	.077	.435	0.060
8. Medications	N/A	+	.952	.077	.603	0.188
9. Poor nutrition	N/A	N/A	.500	.763	.718	4.342 *
10. Feeding tube	N/A	+	.364	.947	.796	9.703 *
11. Need suction	N/A	+	.048	1.000	.762	1.889
12. Poor oral Hyg	N/A	−	.273	.789	.577	0.301
13. DcrMentalSt	N/A	−	.091	.789	.133	1.435
14. Dcr LOC	N/A	+	−	−	−	15
15. COPD	N/A	+	.227	.816	.561	0.161

N/A= not analyzed.
+/ − under ""INTER" indicates whether the reliability kappa for the measure was significant ($P < .05$). Bold indicates a measure is reliable, has sensitivity, and specificity values of at least .600, an A^t of at least .700, and a significant chi-square ($P < .05$).
GI = gastrointestinal; Hyg = hygiene; Dcr = decreased; St = status; LOC = level of consciousness; COPD = chronic obstructive pulmonary disease.
* Indicates the chi-square was significant ($P < .05$).

3.3. Detecting aspiration

The sensitivity and specificity of each C/B E item for detecting thin liquid aspiration are located under the columns ""SENS" and ""SPEC" in Tables 3 - 6. A^t is located under the column labeled ""A^t", and chi-square values are located under the column labeled ""x^2".

3.3.1. History items

None of the History items (see Table 3) met the criteria for sensitivity, specificity, A^t, and chi-square, but three measures were noteworthy for their A^t and chi-square values: a current/history of pneumonia, poor nutrition, and the presence of a feeding tube.

3.3.2. Oral motor signs

None of the oral motor/speech praxis signs (see Table 4) met the stated criteria for sensitivity, specificity, A^t, and chi-square, but three signs were noteworthy for their A^t and chi-square values: the presence or absence of dysarthria, the presence or absence of decreased speech intelligibility, and management of secretions. Volitional cough strength approached criteria in all areas, as well.

Table 4
Reliability, sensitivity, specificity, A^t, and chi-square for oral motor components of a clinical/bedside exam for detecting aspiration

Oral motor	INTRA	INTER	SENS	SPEC	A^t	x^2
1. Tongue						
(a) Strength	+	+	.500	.737	.699	3.436
(b) ROM	+	+	.364	.711	.577	0.354
2. Lips						
(a) Strength (seal)	+	+	.273	.763	.545	0.096
(b) ROM (pucker)	+	+	.476	.718	.670	2.261
(c) ROM (retract)	+	−	.762	.410	.661	1.778
3. Jaw						
(a) Strength	+	+	−	−	−	−
(b) ROM (lateral)	+	+	.381	.744	.624	1.008
4. Soft palate						
(a) Movement	+	+	.409	.711	.615	0.897
(b) Symmetry	+	+	.500	.526	.526	0.039
5. Volitional cough						
(a) Strength	+	+	.700	.450	.637	1.250
(b) Quality	+	+	.545	.684	.689	3.062
6. Palatal gag						
Left	−	+	.714	.154	.240	1.484
Right	−	−	(either side present=present)			
7. Pharyngeal gag						
Left	+	+	.909	.184	.652	0.951
Right	+	+	(either side present = present)			
8. Oral apraxia	+	+	.409	.684	.591	0.533
9. Dysarthria	+	+	.773	.553	.752	6.007 *
10. Intelligibility	+	+	.727	.579	.737	5.249 *
11. Secretions	+	+	.500	.842	.773	8.031 *
12. Attends	+	+	.500	.526	.526	0.039

+/− under ""INTRA" and ""INTER" indicates whether the reliability kappa for the measure was significant ($P<.05$).
Bold indicates a measure is reliable, has sensitivity and specificity values of at least .600, an A^t of at least .700, and a significant chi-square ($P<.05$).
ROM = range of motion.
* Indicates the chi-square was significant ($P<.05$).

3.3.3. Voice signs

Voice quality was rated on two tasks: a speech sample and a sustained /a/. Ratings for each were made and reported separately. None of the voice signs (see Table 5) met the criteria for sensitivity and specificity for detecting aspiration. The most sensitive measure (1.000) for detecting aspiration was an overall rating of the presence of dysphonia as judged by a sustained /a/. Despite its low specificity (.270), the measure had an A^t of .818 and a significant chi-square. The presence or absence of a wet vocal quality was more balanced in regard to sensitivity and specificity and produced A^t values over .700 and significant chi-squares for both sustained /a/ and speech tasks.

Table 5
Reliability, sensitivity, specificity, A^t, and chi-square for voice components of a clinical/bedside exam (obtained from a speech sample and sustained ""AH") for detecting aspiration

	INTRA	INTER	SENS	SPEC	A^t	χ^2
Voice (speech)						
1. Dysphonia	+	+	.909	.216	.679	1.539
2. Breathy	+	+	.818	.297	.632	0.970
3. Harsh	+	+	.818	.297	.632	0.970
4. Strained	−	+	.364	.838	.699	3.094
5. Wet	+	+	.500	.784	.732	5.089 *
6. Resonance	+	+	.455	.811	.728	4.736 *
Voice (AH)						
1. Dysphonia	+	+	1.000	.270	.818	7.159 *
2. Breathy	+	+	.810	.368	.676	2.020
3. Harsh	−	+	.818	.243	.582	0.303
4. Strained	−	−	.524	.737	.713	4.008
5. Wet	+	+	.500	.838	.770	7.677
6. Resonance	+	−	.182	.919	.666	1.339

+/− under ""INTRA" and ""INTER" indicates whether the reliability kappa for the measure was significant ($P < .05$).
Bold indicates a measure is reliable, has sensitivity and specificity values of at least .600, an A^t of at least .700, and a significant chi-square ($P < .05$).
* Indicates the chi-square was significant ($P < .05$).

3.3.4. Signs from trial swallows

Two signs rated on trial swallows met the criteria for sensitivity and specificity: the presence of a spontaneous cough during the swallow, and an overall estimate of the presence or absence of aspiration. Three other items received high A^t values and significant chi-squares: laryngeal elevation, coughing on the 3-oz swallow test, and an overall rating of the presence or absence of dysphagia. That rating was made at the end of the entire C/B E and could be influenced by any other items on the examination.

3.3.5. Detecting aspiration of thicker consistencies

Other individual C/B E signs appeared significant for specific consistencies, but the number of patients who aspirated on consistencies other than thin liquid was too small to provide adequate power. For example, the history sign of poor nutrition was sensitive (.692) and specific (.760) for detecting the presence or absence of thick liquid aspiration. The measure had an A^t of .812 and a significant χ^2 of 7.323, attesting further to its potential merit as an indicator of aspiration on thick liquid. The oral motor sign of volitional cough quality (rated as wet versus dry), was only mildly sensitive (.500) but more specific (.784) for detecting aspiration of puree or solid consistencies. Its A^t was .732, and its χ^2 of 4.268 was significant at $P < .05$. The number of patients who aspirated puree or solid consistencies was too small to provide

Table 6
Reliability, sensitivity, specificity, A^t, and chi-square for trial swallows (thin and thick liquid, puree, and cookie) in a clinical/bedside exam for detecting aspiration

Trial swallows	INTRA	INTER	SENS	SPEC	A^t	x^2
1. Delayed swallow	+	+	.478	.676	.637	1.422
2. TSD	+	−	.304	.811	.630	1.051
3. Laryngeal elevation	+	−	.409	.842	.728	4.689
4. Spontaneous cough	+	+	.682	.816	.835	14.857 *
5. Swallows per bolus	−	+	.091	.895	.457	0.032
6. Wet voice	+	+	.500	.632	.618	0.992
7. Estimate P/A	+	+	.773	.632	.791	9.121 *
8. Oral stasis	−	+	.182	.947	.712	2.584
9. 3-oz swallow	+	+	.864	.500	.787	7.934 *
10. Overall rating of dysphagia	+	+	.909	.474	.807	9.187 *

+/− under ""INTRA" and ""INTER" indicates whether the reliability kappa for the measure was significant ($P<.05$).
Bold indicates a measure is reliable, has sensitivity and specificity values of at least .600, an A^t of at least .700, and a significant chi-square ($P<.05$).
TSD = total swallow duration; P/A = penetration/aspiration.
* Indicates the chi-square was significant ($P<.05$).

adequate power, but a measure of volitional cough quality may be beneficial for predicting aspiration of those consistencies.

4. Discussion

This investigation's purpose was to identify clinical predictors of aspiration on VFSE. It is to be emphasized, however, that predicting such aspiration is but one purpose of the C/B E. In practice, the C/B E serves many other purposes, some of which include the documentation of: feeding position, amount of oral intake, eating efficiency (time to consume a meal), necessity for adaptive feeding equipment, and overall pleasure derived from eating and drinking. Nor does our research emphasis mean that we consider the C/B E a screening examination. As defined by Nielsen and Lang (1999), a screening examination is performed on asymptomatic patients in the search for subclinical disease. The C/B E has a wider range of purposes. One of the most critical, however, is to determine the likelihood that an individual dysphagic patient is aspirating. In present dysphagia practice, such a likelihood is often the motivation for an instrumented examination, usually the VFSE. Sparing facilities and third-party payers unnecessary expense and patients unnecessary radiation exposure associated with the VFSE is important.

Results of this study reveal two items that are reliable, sensitive, specific, and have acceptable A^t and chi-square values for detecting aspiration. These are: the presence of a spontaneous cough during test swallows, and an overall estimate of the presence of aspiration. The sensitivity and specificity of a spontaneous cough

while swallowing test boluses has been identified by others (Daniels et al., 1998; Logemann et al., 1999) However, this is the first study to link a global judgement of the likelihood of aspiration to the presence of aspiration on VFSE.

The importance of spontaneous cough during the swallow is predictable and easily explained. It would only have been surprising if that sign had not emerged from the analysis. To avoid overprediction of aspiration on VFSE, however, the test boluses might better be 5 cc than 3 oz as the 3-oz water test has high sensitivity but modestly low specificity (see Table 6). On the other hand, if one is more concerned with sensitivity than specificity, the 3-oz water test may be appropriate.

More interesting is the sensitivity and specificity of a global judgement about the likelihood of aspiration. Cough during the swallow would, of course, have led a clinician to predict aspiration using the more global item. However, not all aspirators in this study coughed. For those who did not, judges had to rely on other signs. Likely candidates include the presence of a wet voice after the swallow and reduced laryngeal elevation. Moderate sensitivity and specificity have been reported for both (Logemann et al., 1999). Both were moderately sensitive and specific in the present investigation as well.

The literature also provides other candidates. Dysphonia may be the strongest (Daniels et al., 1998; Daniels, McAdam, Brailey, & Foundas, 1997; Horner et al., 1988; Perlman, Booth, & Grayhack, 1994). Based on our findings, the presence of dysphonia is a much stronger indicator of the presence of aspiration than the absence of dysphonia is for indicating the absence of aspiration (see Table 5). In other words, sensitivity (1.000) is remarkably better than specificity (.270), minimizing its utility as a single detector of aspiration versus no aspiration. The presence or absence of a wet vocal quality after swallowing (see Table 5) appears to be related to aspiration as well. However, specificity for that measure was much better than sensitivity.

Previous research has also demonstrated a strong correlation between a weak, volitional cough and aspiration (Daniels et al., 1998; Horner et al., 1988). As Table 4 indicates, this measure was not as promising in the current investigation. Its importance, however, may depend on bolus type. It may be more sensitive and specific for puree and solid consistencies; however, more patients who aspirate puree and solid consistencies would need to be evaluated in subsequent research.

Speech symptoms may also influence a clinician's global judgement of a patient's risk of aspiration. Recent research (Daniels et al., 1997, 1998; Logemann et al., 1999) has provided evidence for the presence of dysarthria as a clinical indicator of aspiration in patients who have suffered a stroke. In the present study dysarthria and speech intelligibility received high A^t values and significant chi-squares (see Table 4). Low specificity decreases their value, however. Many dysarthric patients do not aspirate on VFSE.

Items from the history may also be used by judges in arriving at a global judgement about the likelihood of aspiration on VFSE. For example, previous

research (Splaingard et al., 1988) reported feeding tubes to be the best indicator of aspiration. We found a similar relationship between the two. Similarly, in the present study, the presence or absence of pneumonia or a history of pneumonia and poor nutritional status had high A^t values and significant chi-squares. Their sensitivity for independently detecting aspiration was low, however.

This and other investigations have attempted to identify individual clinical predictors of aspiration (Daniels et al., 1998; Logemann et al., 1999) with variable success. Our minimal criteria for reliability, sensitivity, specificity, A^t, and chi-square, however, eliminated all but two items, one of which is irrelevant for detecting silent aspiration (cough with swallow). Though sensitivity without specificity, and vice versa, may be useful in some situations — depending on whether it is more important to identify disease or save money, under current health care management both the identification of disease and cost are important issues. It would not appear that any single measure is sufficient for predicting aspiration in acute stroke patients.

Another approach is to assemble a combination of reliable clinical signs. Daniels et al. (1998), for example, examined a combination of six measures (dysphonia, dysarthria, abnormal gag, abnormal volitional cough, cough with swallow, and voice change [wet voice] after swallow). The presence of two of any of those six was reported to be more sensitive (.696) and specific (.844) than any single measure for detecting aspiration. Based on those positive results, we examined the sensitivity and specificity of same six signs collectively. Daniels et al. and the current investigation were similar in sample sizes (current = 60; Daniels et al. = 59), and data for both investigations were collected largely from one VA Medical Center) on consecutive patients rather than consecutive referrals. Additionally, each of the six measures was rated with good inter- and intrajudge reliability, without training, in our investigation, indicating that judgements of the clinical features should be replicable in clinical environments. Based on the similar nature of the investigations, we believe such a comparison is worthwhile.

Our individual analyses of sensitivity and specificity were not as impressive as those of Daniels et al. (1998). (See Table 7). Only three of the features had a significant chi-square with our data (as opposed to all six in Daniels et al.), and only the feature of cough with swallow or spontaneous coughing met our sensitivity and specificity criterion. Nevertheless, when utilizing the Daniels et al. criterion of two of six clinical features present for detecting aspiration, both investigations report significant chi-squares. Unfortunately, specificity for that analysis was low for our data (.265). In addition, we looked at sensitivity and specificity using a criterion of four of six features present. Using that criterion, sensitivity (.778) and specificity (.719) for detecting the presence or absence of aspiration were high (A^t was .832), and chi-square was significant (14.446).

As indicated, these investigations were comparable in several ways. However, variability between the two studies probably exists in ratings of individual

Table 7
Six clinical indicators of aspiration from Daniels et al. (1998) and current results

Measure	SENS	SPEC	x^2
Daniels et al., 1998			
Dysphonia	.762	.676	9.98
Dysarthria	.762	.529	4.53
Abnormal gag	.619	.824	11.25
Abnormal voluntary cough	.476	.941	13.25
Cough with swallow	.571	.853	10.95
Voice change after swallow	.381	.853	3.93
Current results			
Dysphonia	1.000	.270	7.159
Dysarthria	.773	.553	6.007
Abnormal gag	.909	.184	0.951
Abnormal voluntary cough	.700	.450	1.250
Cough with swallow	.682	.816	14.857
Voice change after swallow (wet voice)	.500	.632	0.992

Bold indicates chi-square is significant at $P < .05$ or better.

clinical measures, patient severity, and swallowing function within and across patients. Additionally, length of time post-onset could have influenced the results. While patients in both studies were acute, all patients in the Daniels et al. (1998) study were tested within 5 days of admission. Our mean number of days post-onset was 5.98, but patients were included for up to 42 days. Regardless, the two investigations, with separate samples of patients who suffered acute strokes, produced similar, overall results.

Results of the present study confirm that more work needs to be done if C/B E data are to be strongly predictive of aspiration on VFSE. The problems of reliability, sensitivity, and specificity mean that if this form of the C/B E were to be employed, many aspirators would be missed and therefore not referred, and some nonaspirators would be referred unnecessarily, for instrumented examination. As a result, the search for sensitive, specific predictors of aspiration should continue. Combining signs may help, as may asking judges to identify the variables that go into global judgements about risk for aspiration. Methods for improving reliability are also critical because nothing is to be gained from unreliable measurement. In addition, data analysis should be expanded to include negative and positive likelihood ratios (Friedland et al., 1998), which are less dependent on the frequency of aspiration in the experimental sample of patients being tested and may be applied to an array of pretest probabilities based on research and clinical knowledge. Finally it should be recalled that the purpose of the C/B E is not merely to predict aspiration on the VFSE. Forms of the C/B E are utilized worldwide, not only for identifying patients who may aspirate, but also for

defining their overall strengths, limitations, and therapeutic needs for safe, efficient, and pleasurable oral intake. These uses also demand psychometrically sound testing.

Acknowledgments

This work was funded in part by a grant from the Department of Veterans Affairs, Office of Academic Affiliations and a grant from the Department of Veterans Affairs Rehabilitation Research and Development Service.

Appendix A. Continuing education

1. The purpose of a clinical/bedside examination of swallowing is to:
   (a) detect aspiration
   (b) document feeding position
   (c) document eating efficiency
   (d) document need for adaptive eating equipment
   (e) all of the above

2. Which of the following are clinical signs that have been reported in more than one study to be predictive of aspiration in stroke patients:
   (a) dysphonia
   (b) dysarthria
   (c) tongue strength
   (d) a and b
   (e) all of the above

3. Which of the following sections of a clinical/bedside examination have clinical signs which, according to the current results, are sensitive and specific for predicting aspiration:
   (a) history
   (b) oral motor/speech praxis
   (c) voice
   (d) trial swallows
   (e) none of the above

4. Daniels et al. (1998) report that six clinical signs are the most sensitive and specific for detecting aspiration. How many of those six do they say should be present before a patient is suspected of aspirating:
   (a) one
   (b) two
   (c) three

(d) four
(e) all of them

5. Which of the following do the authors say are still potential problems with clinical/bedside signs for detecting aspiration:
   (a) reliability of clinical signs
   (b) unnecessary referrals for VFSE
   (c) radiation exposure
   (d) a and b
   (e) all of the above

References

Arvedson, J., Rogers, B., Buck, G., Smart, P., & Msall, M. (1994). Silent aspiration prominent in children with dysphagia. *International Journal of Pediatric Otorhinolaryngology*, *28*, 173-181.

Chial, M. R. (1997). *Clinical decision analysis: manual notes*. Unpublished.

Chial, M. R. (1998). Personal communication.

Chokshi, S., Asper, R., & Khandheria, B. (1986). Aspiration pneumonia: a review. *American Family Physician*, *33* (3), 195-202.

Cogen, R., & Weinryb, J. (1989). Aspiration pneumonia in nursing home patients fed via gastrostomy tubes. *American Journal of Gastroenterology*, *84* (12), 1509-1512.

Daniels, S. K., Brailey, K., Priestly, D. H., Herrington, L. R., Weisberg, L. A., & Foundas, A. L. (1998). Aspiration in patients with acute stroke. *Archives of Physical Medicine and Rehabilitation*, *79*, 14-19.

Daniels, S. K., McAdam, C. P., Brailey, K., & Foundas, A. L. (1997). Clinical assessment of swallowing and prediction of dysphagia severity. *American Journal of Speech-Language Pathology*, *6*, 17-24.

DeJong, R. (1967). *The neurologic examination*. New York: Hoeber Medical Division, Harper & Row.

DePippo, K. L., Holas, M. A., & Reding, M. J. (1992). Validation of the 3-oz water swallow test for aspiration following stroke. *Archives of Neurology*, *49*, 1259-1261.

Ekberg, O., Nylander, G., Fork, F. T., Sjoberg, S., Birch-Iensen, M., & Hillarp, B. (1988). Interobserver variability in cineradiographic assessment of pharyngeal function during swallow. *Dysphagia*, *3*, 46-48.

Feinberg, M. J., Ekberg, O., Segall, L., & Tully, M. A. (1992). Deglutition in elderly patients with dementia: findings of videofluorographic evaluation and impact on staging and management. *Radiology*, *183*, 811-814.

Friedland, D. J., Go, A. S., Davoren, J. B., Shlipak, M. G., Bent, S. W., Subak, L. L., & Mendelson, T. (1998). *Evidence-based medicine: a framework for clinical practice*. Stamford, CT: Appleton & Lange.

Garon, B. R., Engle, M., & Ormiston, C. (1996). Silent aspiration: results of 1,000 videofluoroscopic examinations. *Journal of Neurogenic Rehabilitation*, *10*, 121-126.

Gordon, C., Hewer, R. L., & Wade, D. T. (1987). Dysphagia in acute stroke. *British Medical Journal*, *295*, 411-414.

Hartelius, L., & Svensson, P. (1994). Speech and swallowing symptoms associated with Parkinson's disease and multiple sclerosis: a survey. *Folia Phoniatrica Logopaedia*, *46*, 9-17.

Horner, J., Brazer, S. R., & Massey, E. W. (1993). Aspiration in bilateral stroke patients: a validation study. *Neurology*, *43*, 430-433.

Horner, J., Massey, E. W., & Brazer, S. R. (1990). Aspiration in bilateral stroke patients. *Neurology*, *40*, 1686-1688.

Horner, J., Massey, E. W., Riski, J. E., Lathrop, D. L., & Chase, K. N. (1988). Aspiration following stroke: clinical correlates and outcome. *Neurology*, *38*, 1359-1362.

Kuhlemeier, K. V., Yates, P., & Palmer, J. B. (1998). Intra- and interrater variation in the evaluation of videofluorographic swallowing studies. *Dysphagia*, *13*, 142-147.

Leder, S. B. (1997). Videofluoroscopic evaluation of aspiration with visual examination of the gag reflex and velar movement. *Dysphagia*, *12*, 1221-1223.

Leder, S. B., Sasaki, C. T., & Burrell, M. I. (1998). Fiberoptic endoscopic evaluation of dysphagia to identify silent aspiration. *Dysphagia*, *13*, 19-21.

Linden, P., Kuhlemeier, K. V., & Patterson, C. (1993). The probability of correctly predicting subglottic penetration from clinical observations. *Dysphagia*, *8*, 170-179.

Linden, P., & Siebens, A. A. (1983). Dysphagia: predicting laryngeal penetration. *Archives of Physical Medicine and Rehabilitation*, *64*, 281-284.

Logemann, J. A. (1998). *Evaluation and treatment of swallowing disorders* (2nd ed.). Austin, TX: Pro-Ed.

Logemann, J. A., Veis, S., & Colangelo, L. (1999). A screening procedure for oropharyngeal dysphagia. *Dysphagia*, *14*, 44-51.

McCullough, G. M., Wertz, R. T., Rosenbek, J. C., & Dinneen, C. (1999). Clinicians' preferences and practices in conducting clinical/bedside and videofluoroscopic examinations of swallowing. *American Journal of Speech-Language Pathology*, *8*, 149-163.

McCullough, G. M., Wertz, R. T., Rosenbek, J. C., Mills, R. H., Ross, K., & Ashford, J. (2000). Inter- and intrajudge reliability of clinical/bedside examination measures of swallowing in adults. *Dysphagia*, *15*, 58-67.

McCullough, G. M., Wertz, R. T., Rosenbek, J. C., Mills, R. H., Ross, K., & Webb, W.G. (in press). Inter- and intrajudge reliability of videofluoroscopic measures of swallowing in adults subsequent to stroke. *Dysphagia*.

Nielsen, C., & Lang, R. S. (1999). Principles of screening. *Medical Clinics of North America*, *83* (6), 1323-1337.

Perlman, A. L., Booth, B. M., & Grayhack, J. P. (1994). Videofluoroscopic predictors of aspiration in patients with oropharyngeal dysphagia. *Dysphagia*, *9*, 90-95.

Scott, A., Perry, A., & Bench, J. (1998). A study of interrater reliability when using videofluoroscopy as an assessment of swallowing. *Dysphagia*, *13*, 223-227.

Smith, C. H., Logemann, J. A., Colangelo, L. A., Rademaker, A. W., & Pauloski, B. R. (1999). Incidence and patient characteristics associated with silent aspiration in the acute care setting. *Dysphagia*, *14*, 1-7.

Splaingard, M. L., Hutchins, B., Sulton, L. D., & Chaudhuri, G. (1988). Aspiration in rehabilitation patients: videofluoroscopic versus bedside clinical assessment. *Archives of Physical Medical Rehabilitation*, *69*, 637-640.

Wilcox, F., Liss, J. M., & Siegel, G. M. (1996). Inter-judge agreement in videofluoroscopic studies of swallowing. *Journal of Speech and Hearing Research*, *39*, 144-152.