Appalachian
STATE UNIVERSITY.
BOONE, NORTH CAROLINA

Utility Of Clinical Swallowing Examination Measures For Detecting Aspiration Post-Stroke

Authors
McCullough, G.H., Rosenbek, J.C., Wertz, R.T., McCoy, S., Mann, G., & McCullough,K.

Abstract

The purpose of this investigation was to determine the utility of clinical swallowing examination (CSE) measures for detecting aspiration as defined by videofluoroscopic swallowing examination (VFSE). This study, involving 165 participants, is a follow-up to a previously published investigation of 60 participants. Findings are compared with that investigation as well as with other research on CSEs. The results suggest that clinicians can make an accurate judgment of the occurrence of aspiration in most poststroke patients. However, ruling out aspiration when it is absent appears more problematic. More work needs to be done if data collected from noninstrumented examinations are to be strongly predictive of the presence and absence of aspiration on VFSE. At present, there are no data to suggest that CSEs can be used to quantify aspiration or make adequate recommendations regarding patient care.

Utility of Clinical Swallowing Examination Measures for Detecting Aspiration Post-Stroke

Evaluations of swallowing function can be conducted using various methodologies depending on the stage(s) of the swallow one needs to assess and on the clinician's or experimenter's purpose. Speech–language pathologists attempting to evaluate oral, pharyngeal, and cervical esophageal function in medical settings most frequently choose one (and often two) of three options: (a) clinical swallowing examination (CSE), (b) videofluoroscopic swallowing examination (VFSE), or (c) fiberoptic endoscopic examination of swallowing (FEES). Each examination has strengths and weaknesses, and data to define those strengths and weaknesses are continually emerging. VFSE has, historically, been considered the gold standard, largely for its utility in assessing oral, pharyngeal, and cervical esophageal stages of swallowing. However, it is often difficult to obtain because of transport and other issues with patients, and it is necessarily brief because of the use of radiation. Furthermore, it is unnatural, because it examines swallowing function in idealized circumstances, with upright posturing and coaching, and uses boluses that only loosely approximate normal food and liquid intake. FEES, on the other hand, can be an excellent tool for assessing numerous aspects of the pharyngeal swallow over time in a more natural feeding environment with a variety of foods but does not allow for a thorough assessment of oral or cervical–esophageal function (Daniels et al., 2003). A CSE, on the other hand, provides no opportunity to directly observe the physiology of the swallow at any stage. Nevertheless, pertinent historical information and information regarding oral motor and feeding abilities can be gathered.

Furthermore, VFSE and FEES are costly examinations, requiring expensive equipment that most clinicians in home health and nursing home settings do not possess. Thus, there is a need to validate what a CSE assesses accurately and adequately.

Martino, Pron, and Diamant (2000) reviewed the literature on CSEs extensively, evaluating 154 sources, 89 of which were original articles. Data, when available, were collapsed and reanalyzed for sensitivity, specificity, and likelihood ratio. Their results suggested that few data are currently available to support the concept that clinicians are able to detect abnormal swallow physiology with a clinical examination, and they suggested that "large, well-designed trials are needed for more conclusive evidence of screening benefit" (Martino et al., 2000, p. 19). The concerns expressed by Martino et al. are understandable when one examines some of the individual clinical signs reported in the literature. The presence of 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, Massey, Riski, Lathrop, & Chase, 1988; Linden & Siebens, 1983; Logemann, Veis, & Colangelo, 1999) have been identified by some researchers as signs of aspiration in stroke patients. Others have found no relationship between an abnormal, volitional cough or the lack of a pharyngeal gag reflex and aspiration (Leder, 1997; Linden, Kuhlemeier, & Patterson, 1993).

Other signs that have been investigated 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., 1993; Horner et al., 1990; Horner et al., 1988; 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).

Collections of signs have also been investigated. Daniels et al. (1998) reported that the presence of any two of six clinical signs (dysphonia, dysarthria, abnormal gag, abnormal volitional cough, cough with swallow, and voice change—wet voice—after swallow) is highly predictive of aspiration when compared with videofluoroscopy. Leder and Espinosa (2002) studied the same six clinical signs in comparison with FEES and reported underestimation in patients with aspiration and overestimation in patients who did not aspirate. Reliability for rating the clinical signs was not reported in either study. This may be one reason for variability, as well as differences in the gold standard used (VFSE vs. FEES). Daniels, McAdam, Colleen, Brailey, and Foundas (1997) further reported that the same two of six signs were predictive of dysphagia severity.

It is difficult to know what to conclude from the controversies surrounding the reports for clinical signs of aspiration and dysphagia. 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. Sensitivity and specificity may differ depending on whether patients are enrolled consecutively or selectively and whether they are referred by another health care 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 and signs for the CSE. Inter- and intrajudge reliability data for rating CSE signs have been reported in only two studies (McCullough, Wertz, & Rosenbek, 2001; McCullough et al., 2000). The assumption appears to be that judges can reliably evaluate responses to CSE measures, although evidence of reliability problems for these measures was reported in both of those studies. Finally, limited control over the kinds and viscosities of boluses swallowed at bedside and during VFSE seems to betray an assumption that such control is not critical.

The purpose of this investigation was to further examine the utility of CSE measures and signs for detecting aspiration. Results from a previous study using the same bolus volumes and viscosities, the same clinical signs, and the same enrollment criteria are available for comparison (McCullough et al., 2001). Comparisons without those assumptions being met have not previously been reported. It must be emphasized that we do not believe aspiration is the only important factor in assessing for dysphagia; neither do we believe that detecting aspiration clinically can define swallow physiology or make appropriate dietary and treatment recommendations. We seek to provide evidence, positive or negative, regarding the utility of CSE signs—measured in a consistent, reliable, replicable way—for detecting aspiration.

Our primary research question was: Do individual CSE measures or combinations of CSE measures detect aspiration in patients who have suffered a stroke? We hypothesized that clinicians with experience can make a global judgment of aspiration but that the nature of the

judgment may be more global (based on the examination as a whole) than specific to individual measures.

Method
Participants

One hundred sixty-five patients who had suffered an acute, ischemic stroke and agreed to participate in the investigation were consecutively enrolled. Participants were recruited from the Veteran's Affairs (VA) Medical Centers in Nashville, Tennessee, and Madison, Wisconsin. Inclusion criteria for participation in the study were occurrence of a stroke within 6 weeks of the time of examination (over 86% were within 1 week poststroke, and 95% were within 2 weeks poststroke) and competence to provide informed consent. Exclusion criteria were presence of a structural anomaly that could interfere with swallowing, presence or recent history of a tracheostomy, a reported history of dysphagia prior to the stroke, and/or physician judgment that the patient was too medically unstable to participate in the study. Individuals who had experienced previous strokes were allowed to participate as long as no swallowing problems resulted from those strokes.

One hundred sixty-three of the 165 participants were male. The mean age of the participants was 65 years, with a range of 39 to 101 years and a standard deviation of 11 years. One hundred fifty were right handed, 14 were left handed, and 1 individual was ambidextrous (according to patient or family report). The majority of participants (140) were Caucasian, 20 were African American, 2 were American Indian, and race was not determined for 3.

CSE

After informed consent was obtained, the primary study clinician performed a CSE on each participant. There were four sections to the CSE: (a) history, (b) oral motor, (c) voice and speech praxis, and (d) trial swallows. Clinical signs were rated using a binary system (+/−) to indicate whether the clinical sign was or was not observed. Selection of clinical signs was based on reports of clinicians' preferences and practices for evaluating dysphagia in adults (McCullough, Wertz, Rosenbek, & Dinneen, 1999) and was consistent with the previous investigation (McCullough et al., 2001). 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. Mental status was established using the Modified Mini-Mental State Examination (Bravo & Hebert, 1997), and a score of 59 or below resulted in a positive rating for decreased mental status. The Barthel Index assesses patients for activities of daily living and has been shown to be a valid, reliable tool for research and clinical practice (Wade & Collin, 1988). The Barthel Index was, therefore, used as an indicator of normal or abnormal functional status. Scores can range from 0 to 20, where 20 = independent, 15–19 = mild impairment, 10–14 = moderate impairment, 5–9 = severe impairment, and 0–4 = very severe impairment. Participants are rated on feeding, bathing, grooming, dressing, bladder control, bowel control, toileting, chair and bed transfer, mobility, and stair climbing. These scores were obtained by the speech–language pathologist doing the clinical examination in conjunction with information obtained from physical therapists, occupational therapists, and nurses. Patients with moderate to very severe scores (0–14) were considered to have decreased functional status. Nutritional status was derived from the medical chart based on the evaluation of a registered dietician. Dieticians rated the patients' nutritional status as normal, mildly compromised, moderately compromised, or severely compromised based on a complex array of factors, including diagnoses indicative of nutritional risk (i.e., alcoholism, cancer, etc.), physical findings (i.e., cachexia, weight loss, poor dentition, etc.), hospital treatments (i.e., medications, prolonged inadequate diet, nothing by mouth for extended period, etc.), and psychosocial factors (i.e., fear, anxiety, etc.). Laboratory values also play an important role; the most useful for making an overall statement about nutritional status, according to our dieticians, was albumin. Albumin has also been demonstrated to improve with treatment for dysphagia (Elmstahl, Bulow, Ekberg, Petersson, & Tegner, 1999). "Poor oral hygiene" was a judgment made by the clinician. Although few guidelines were available for such a judgment, we attempted to make it on the basis of overall cleanliness of the oral cavity, appearance of the teeth or dentures, and appearance of any oral diseases or tooth decay. Missing teeth would not produce a negative rating here unless oral disease or uncleanliness was apparent.

Oral motor measures primarily related to the strength and appearance of the structures involved in oral preparation and bolus propulsion. Structures were observed for symmetry as well as tone. Tongue strength was measured with anterior, lateral, superior, and inferior movement with resistance by a tongue blade. Lips strength was measured by having the participant puff up his cheeks with air while the clinician pressed against the cheeks and by having the participant purse his lips while the clinician tried to separate them. Lip protrusion and retraction were also examined. Jaw strength was assessed by opening and closing the jaw against resistance with the clinician's hand. The palatal and pharyngeal gags were assessed using cotton-tip applicators applied to both the left and right sides of those structures and observing for a response. Patients were also asked to cough, volitionally, and were rated on

the strength of that cough as well as the quality (wet or dry sounding). The reflexive cough was rated the same way but was assessed only if the participant exhibited such a cough during the testing.

Voice and speech measures were elicited by having the participant read "The Grandfather Passage" or describe the picture from the Western Aphasia Battery (Kertesz, 1982). The voice was rated perceptually first for any type of dysphonia and then, more specifically, for wet/gurgly quality, breathiness, or strained/strangled quality. Nasal resonance was assessed perceptually for hyper- or hyponasality. The measurements of aphasia, dysarthria, voice, and apraxia of speech were all made while listening to the reading or description. Oral apraxia was determined by having the participant cough, click the tongue, blow, bite the lower lip, and puff out the cheeks.

For the trial swallows portion of the CSE, two swallows of each consistency—5 ml thin liquid, 10 ml thin liquid, thick liquid, puree, and solid (1/4 cookie) were administered. Thin and thick liquids were administered from a pill cup; puree and solids were administered from a spoon. Finally, when it was deemed safe and appropriate, the participant was tested with the 3-oz swallow test (DePippo et al., 1992). "Safe and appropriate" means that the clinician made a judgment based on the rest of the examination (including history, oral motor, speech/praxis, and trial swallows) that the patient had not been placed at risk by the prior trial swallows and/or history or current status. If the patient had already exhibited signs of moderate to severe impairment of swallowing and was considered to have compromised medical status based on related diagnoses, the 3-oz swallow test was not used. This was a judgment made by the clinician.

For all swallows, the clinician elicited pre- and postswallow voicing ("Ah") and used the four-finger method of laryngeal palpation (Logemann, 1998). Postswallow voicing was judged for a wet/gurgly voice quality in comparison with baseline (preswallow voicing). The four-finger method was used to determine the timing and completeness of the swallow as well as the number of swallows necessary for bolus passage. As Logemann (1998) described, the initiation of bolus propulsion is felt by initial submental contraction of the mylohyoid. This movement is considered in relation to the initiation of hyoid and laryngeal movement for measures such as "delayed oral transit" and delayed swallow onset (such as stage transition duration). Multiple swallows (more than two for any bolus) may be considered a sign of some type of oral or pharyngeal weakness limiting bolus propulsion. Clinicians, therefore, rated whether they believed the participant displayed a delayed oral transit as well as whether the participant experienced laryngeal penetration or aspiration. In addition, the clinicians listed the reasons they suspected penetration/aspiration, such as

the presence of coughing or clearing of the throat, wet voice quality, change in respiration, watering eyes, or the participant's report of a problem.

VFSE

Within 24 hr after the completion of the CSE, another speech pathologist, blind to the results of the CSE, administered a VFSE. Participants were seated upright in a wheelchair or stretcher chair for the duration of the study. At the Nashville VA center, studies were conducted with a mobile C-arm X-ray (Model 9400) system (OEC-Diagnostics, Inc.), 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 Madison VA, the studies were conducted with a Siemens fluoro unit (Model 8842437G5275) with a 40-in. (101.6-cm) fixed tower and recorded with a Panasonic AG6300 MD videocassette recorder.

Each examination began with two 5-ml then two 10-ml thin liquid swallows (50/50 mixture of water and E-Z-HD Barium Sulfate Powder for Suspension @ 14 centipoise). Next, the participant swallowed two 5-ml boluses of thick liquid (thickened juice and barium powder @ 187 centipoise). These swallows were followed by two 5-ml boluses of applesauce (mixed with barium powder) followed by two solids (1/4 Lorna Doone cookie coated with Barium Sulfate Esophageal Cream). Finally, when deemed safe and appropriate, participants were given a cup with 3 oz (89 ml) of thin liquid barium and instructed to drink it as quickly as possible, using consecutive, uninterrupted swallows. All swallows were viewed in the lateral plane. At least 1 week after the completion of the videofluoroscopic examination, and when at least five study tapes had been collected, a study clinician, blinded to information regarding the participant and his or her CSE, viewed the videotaped videofluoroscopy studies and made her ratings. Only the rating of aspiration as present or absent are discussed in this report. Aspiration was considered to occur when a bolus passed below the level of the true vocal folds.

Reliability

Inter- and intrajudge reliability for both the CSE and VFSE measures were obtained on a random sample of 15% of the participants. For the CSE, interjudge reliability measures were obtained by having a second clinician make ratings and judgments along with the primary study clinician in each participating medical center. Therefore, the primary study clinician and the reliability clinician examined the patient at the same time. The clinicians did not discuss the judgments they made during or after the examination. Intrajudge reliability was

Figure 1. Signal detection 2 x 2 contingency table demonstrating sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio of one clinical/bedside swallowing sign (wet voice after swallow) for detecting aspiration. VFSE = videofluoroscopic swallowing examination; TP = true positive; FP = false positive; FN = false negative; TN = true negative.

	Aspiration on VFSE		
	Positive	Negative	Row Sums
Positive	20 17		37
Dysphonia	a/TP b/FP		
	c/FN d/TN		
Negative	17 10		120
Column Sums	37	120	157 N

```
Sensitivity = a/(a + c) = 20/(20 + 17) = 54%
Specificity = d/(b + d) = 103/(17 + 103) = 86%
Positive Predictive Value = a/(a + b) = 20/(20 + 17) = 54%
Negative Predictive Value = d/(c + d) = 103/(17 + 103) = 86%
Positive Likelihood Ratio = sensitivity/(1 ñ specificity)
                          = .54/(1 ñ .86) = 3.8
```

obtained by having the primary study clinician administer a second CSE to 15% of all participants, again randomly selected, the day after the initial evaluation. Although time can be a factor in the recovery from acute stroke, this was the most appropriate method we could derive for obtaining intrajudge measurements. Even if aspiration was suspected, the CSE was administered to intrajudge participants the next day. The same bailout criteria were used for participant safety. This method was used in a previous investigation (McCullough et al., 2000). Reliability for CSEs was established within, not across, the two study hospitals.

For intrajudge reliability of VFSE measures, the study clinician who performed all VFSE ratings 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. All clinicians who participated in this study had over 200 hr of experience with CSE and VFSE examinations of swallowing.

Statistical Analysis

A signal detection analysis program (Chial, 1997) was used to determine sensitivity, specificity, positive predictive value, negative predictive value, and likelihood ratio

(defined below) of CSE measures for detecting: aspiration and prolonged stage transition duration as determined by the VFSE. A 2 x 2 contingency table was used (see Figure 1) to evaluate each CSE measure. If aspiration was present on VFSE when a history item or sign was present on the CSE, 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 CSE sign was present, a *false positive* rating resulted. If aspiration was present on the VFSE but the CSE sign was absent, a *false negative* rating resulted. Signal detection theory is typically used for determining the above-mentioned values for the purpose of detecting the presence or absence of disease. We used the values in this investigation simply to determine whether a relationship exists.

Sensitivity for each CSE item was computed by dividing the number of participants with a true positive clinical sign by the total number of participants who aspirated on VFSE. Specificity for each CSE item was computed by dividing the number of participants with a true negative clinical sign by the total number of participants who did not aspirate on VFSE. Positive predictive value (PPV) was computed by dividing the number of participants with a true positive clinical sign by the total number of participants who tested positive for that sign. Negative predictive value (NPV) was computed by dividing the number of participants who were negative for that

clinical sign by the total number of people who tested negative for that sign. Therefore, whereas sensitivity and specificity address the proportion of participants who aspirated and were positive or negative for clinical signs, PPV and NPV address the number of people with or without a sign who did and did not aspirate. This distinction is important when examining differences reported in the Results section. A likelihood ratio (LR) is derived by dividing sensitivity by 1 minus the specificity. LR provides a type of odds ratio rather than a percentage. For example, someone with dysphonia in this study was 3.8 times more likely to aspirate than someone without dysphonia. An LR for a clinical sign positively detecting aspiration can be high even if sensitivity is low. This is true because it is high *specificity* that helps *rule in* a diagnosis when a clinical sign is present (Sackett, Richardson, Rosenberg, & Haynes, 1998). High *sensitivity* is more indicative that a negative sign can help *rule out* a diagnosis.

For analysis of inter- and intrajudge reliability of the binary ratings for CSE measures and the VFSE measure of aspiration, we used Cohen's kappa, a chance-corrected measure of agreement that is not based on chi-square but that does use contingency tables (Howell, 1992, pp. 148–150). According to Howell (1992), the significance of kappa is not an issue; rather, if kappa is low enough to even question significance, then agreement is poor. Kappas range from 0 to 1 for standard correlations or from 0 to –1 if there is an inverse correlation.

In addition to calculating sensitivity, specificity, PPV, NPV, and LR for each clinical measure, we selected the clinical measures that were most strongly associated with the dependent measure (aspiration) and submitted them to a backward stepwise multiple logistic regression to generate a prediction model of best fit. Several models were generated and reviewed on the basis of goodness-of-fit statistics and predicted versus observed probabilities. Only those models that performed optimally and met biologic plausibility were accepted.

Results
VFSE Results

Aspiration observed in the VFSE was used as the gold standard to determine the sensitivity, specificity, PPV, NPV, and LR for each clinical measure. Forty-three participants (26%) were judged to aspirate. Twenty-two (51%) were "silent" aspirators. The occurrence of aspiration poststroke was observed more often in the eldest age group (80–101 years = 40% aspirated) than in the younger age groups (39–59 years = 23% aspirated, 60–79 years = 25% aspirated).

Sixty-two participants showed evidence of a prior stroke, but none of those participants had known swal-

Table 1. Lesion localization and the number of patients who did and did not aspirate.

Localization	N	Aspiration	No aspiration	% Aspirated
Cortical				
Right	39	13	26	33
Left	37	8	29	22
Bilateral	6	1	5	17
Subcortical				
Right	23	5	18	22
Left	17	4	13	24
Bilateral	16	7	9	44
Brain stem				
Right	11	2	9	18
Left	10	2	8	20
Bilateral	1	1	0	100
Cerebellar	4	0	4	0
Mixed	1	0	1	0

lowing problems from the prior stroke. Data from brain imaging and neurological examinations, shown in Table 1, indicate that 82 participants had sustained a cortical stroke or strokes. Fifty-six participants had sustained a subcortical stroke or strokes. Twenty-two participants had sustained a brain stem stroke or strokes. Four participants had cerebellar strokes, and 1 had a lesion that could not be specifically localized. Aspiration occurred more frequently subsequent to right hemisphere strokes than left. Although previous research has also observed this relationship (Robbins & Levine, 1988), the relationship in this investigation, although present, did not appear to be as strong and was not observed in subcortical or brain stem lesions. Bilateral subcortical stroke patients had an increased occurrence of aspiration over most other unilateral and bilateral stroke patients (44%). Only 1 patient in our sample had a bilateral brain stem stroke, but that patient did aspirate. Twenty-two (50%) of the 43 aspirators had suffered a prior stroke without prior history of dysphagia.

Reliability

Interjudge reliability for each CSE measure is located in the last column of Tables 2 through 5. Percentage agreement is shown for each, as is Cohen's kappa when it was computable. Kappa could not be computed if one clinician's judgments were the same across all reliability participants or if insufficient data were available to compute the calculation. Percentage agreement ranged from 46% (reduced jaw strength; see Table 3) to 100% (for 11 of the 28 history measures in Table 2 and 8 of the 9 voice and speech praxis measures in Table 4). Cohen's kappa values ranged from –.17 (delayed oral transit with

Table 2. Sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and interjudge reliability (REL) of history signs for detecting aspiration in a clinical swallowing examination.

	SENS	SPEC	PPV	NPV	+LR	REL[a]
Patient report	54	71	37	83	1.8	100 (1.0)
Family report	67	73	48	85	2.4	92
Nurse report	57	80	49	85	2.8	92 (.80)
Nonoral feeding	49	84	51	83	3.1	92 (.72)
Mental status	46	59	27	77	1.1	100
Barthel Index	79	49	87	58	1.5	100 (1.0)
Pneumonia	9	98	67	76	5.8	100 (1.0)
History pneumonia	5	88	40	34	0.4	77 (.38)
Medications	33	67	26	74	1.0	100 (1.0)
Need suction	5	98	50	75	2.9	92
Poor nutrition	12	90	29	75	1.2	100
Dehydration	7	98	50	75	2.9	100
Drools	23	94	56	78	3.6	92 (.62)
Poor oral hygiene	14	97	60	77	4.4	92 (.70)
Edentulous	35	70	29	76	1.2	100
Dentures	35	69	28	75	1.1	92 (.62)
Alcohol use	24	76	26	75	1.0	100
Hypertension	84	17	26	75	1.0	100 (1.0)
Tobacco use	37	66	28	75	1.1	92 (.61)
Reflux	16	89	33	75	1.4	92 (.62)
Diabetes	30	67	24	74	0.9	92 (.81)
Obese	33	66	25	74	1.0	100 (1.0)
Pulmonary disease	12	85	21	73	0.8	92 (.79)

Note. Unless otherwise noted, all table values are percentages.

[a]Interjudge reliability is reported as percentage agreement and kappa (in parentheses) when computable.

3-oz swallow; see Table 5) to 1.0 (various measures in each section and 8 of the 9 measures in the voice and speech praxis section; see Table 4). Only three measures in the CSE were rated with less than 70% agreement: (a) reduced jaw strength (46%), (b) penetration of 10 ml thin liquid (69%), and (c) penetration of thick liquid (54%). Nonetheless, when percentage agreement dropped below 80%, kappa values dropped substantially—in a couple of measures, even into low and negative (inverse) values. Kappa attempts to adjust percentage agreement to correct for chance agreement. Typically, when percentage agreement for these ratings was 80% or greater, kappa values were within an acceptable range, although the range of acceptability varies depending on one's source. Intrajudge percentage agreement was 80% or greater for all measures; all Cohen's kappas for intrajudge reliability were greater than .40, and 90% were greater than .50. These results are consistent with a prior investigation (McCullough et al., 2000) that suggested individuals with training and experience in administering CSEs can become reliable with themselves but that establishing reliability with other clinicians requires training. That training should only truly take place when measures of great importance are firmly established.

Inter- and intrajudge reliability for rating the presence or absence of aspiration from VFSEs was very high ($k = .965$, $p = .000$).

Detecting Aspiration

The answer to our primary research question, "Do individual CSE measures or combinations of CSE measures detect aspiration in patients who have suffered a stroke?" is provided in Tables 2 through 5. The sensitivity, specificity, PPV, NPV, and LR of each CSE measure for detecting aspiration are located under the columns labeled *SENS, SPEC, PPV, NPV,* and *LR,* respectively, in those tables. Specific results are divided by CSE section below.

History items. As shown in Table 2, the two history measures with the most utility for detecting aspiration (based on LR) are (a) the presence of a pneumonia and (b) poor oral hygiene. Sensitivity is low for both measures, and specificity is high. Recall that high specificity means that a positive sign is very helpful for ruling in aspiration, which is reflected in the higher LR. Low sensitivity, on the other hand, indicates that a negative

Table 3. Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and interjudge reliability of oral motor signs for detecting aspiration in a clinical swallowing examination.

	SENS	SPEC	PPV	NPV	+LR	REL[a]
Tongue: Strength						
Weak (L or R)	64	48	30	79	1.2	90
Weak bilateral	36	80	39	78	1.8	100
Structure abnormal	2	94	12	74	0.4	85 (.58)
Lips: Strength						
Weak (L or R)	68	49	31	82	1.3	77 (.44)
Weak bilateral	17	87	30	76	1.3	85 (.62)
Structure abnormal	51	68	36	80	1.6	100
Jaw: Strength						
Weak (L or R)	26	96	67	79	5.8	46 (−.02)
Weak bilateral	15	99	86	77	17.4	92
Structure abnormal	7	97	43	76	2.2	54 (−.05)
Soft palate: Strength						
Weak (L or R)	47	74	36	82	1.8	77 (.08)
Weak bilateral	25	91	45	78	2.6	91 (.62)
Structure abnormal	24	93	53	79	3.4	100
Weak palatal gag						
L or R	56	51	29	77	1.2	77 (.52)
Bilateral	56	60	33	80	1.4	77 (.40)
Weak pharyngeal gag						
L or R	57	54	29	79	1.2	77 (.25)
Bilateral	54	66	35	81	1.6	85 (.42)
Volitional cough						
Strength	42	79	39	81	2.0	100 (1.0)
Quality	26	89	41	80	2.3	85 (.56)
Reflexive cough						
Strength	24	80	38	66	1.2	85
Quality	38	77	46	70	1.6	92 (.63)

Note. Unless otherwise noted, all table values are percentages. L = left; R = right.

[a]Interjudge reliability is reported as percentage agreement and kappa (when computable).

result is not very helpful for ruling aspiration out. We might expect, therefore, that PPV should be high and NPV should be lower, which is not the case. However, sensitivity and specificity address the proportion of participants who aspirated and were positive or negative for clinical signs, whereas PPV and NPV address the number of people with or without a sign who did and did not aspirate. They are derived in very different manners, which can produce different results. LR is derived from sensitivity and specificity and is, therefore, more congruent with those values. A high LR should, in theory, occur with higher specificity, as a positive sign should make it easier to rule in aspiration. In short, the presence of pneumonia and poor oral hygiene provide much more information than their absence. The presence of nonoral feeding equipment (i.e., patient was tube fed) and drooling are the only other two measures with LRs of 3.0 or greater.

Oral motor signs. As shown in Table 3, the most useful oral motor signs for detecting aspiration relate to jaw function. Unfortunately, these measures are among the worst in terms of interjudge reliability. Bilateral weakness has the highest LR, and unilateral weakness has the second highest LR. Again, based on the low sensitivity, we cannot presume that the absence of jaw weakness indicates an aspiration-free swallow. Only 15 of the 165 participants demonstrated unilateral jaw weakness, and only 7 demonstrated bilateral jaw weakness. Abnormal jaw structure (muscle or bone atrophy/changes) was the only other measure with a likelihood ratio of 3.0 or greater.

Voice/speech/praxis signs. Voice quality was rated on a binary scale using a speech sample. As shown in Table 4, the two best voice measures for detecting aspiration were "breathy voice" (LR = 6.6) and "wet/gurgly voice" (LR = 5.3). Sensitivity was low for both measures, but specificity values were 98% and 96%, respectively. PPVs of 67% and 62% indicate that predicting aspiration from these measures, although helpful, remains suspect. Two other measures—(a) "strained voice" and (b) a

Table 4. Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and interjudge reliability of voice and speech praxis signs for detecting aspiration in a clinical swallowing examination.

	SENS	SPEC	PPV	NPV	+LR	REL[a]
Dysphonia	54	86	54	86	3.8	77 (−.13)
Wet/gurgly voice	22	96	62	80	5.3	100 (1.0)
Breathy voice	16	98	67	79	6.6	100 (1.0)
Strained voice	30	92	52	81	3.6	100 (1.0)
Nasal resonance	44	84	44	84	2.8	100 (1.0)
Dysarthria	78	46	32	86	1.4	100 (1.0)
Oral apraxia	16	91	35	78	1.8	100 (1.0)
Speech apraxia	3	94	13	78	0.5	100 (1.0)
Aphasia	33	78	33	79	1.5	100 (1.0)

Note. Unless otherwise noted, all table values are percentages.

[a]Interjudge reliability is reported as percentage agreement and kappa (when computable).

general rating of "dysphonia,"—also indicated aspiration was at least three times more likely (LR > 3) when the measure or sign was present.

Signs from trial swallows. As shown in Table 5, the best trial swallow measures (according to LR) for detecting aspiration were global judgments of aspiration made from 3 oz thin liquid (LR = 9.5), 10 ml thin liquid (LR = 9.2), 5 ml thick liquid (LR = 8.7), and 5 ml thin liquid (LR = 6.8). All global judgments of aspiration made from trial swallows produced LRs of greater than 4.0, indicating that when aspiration was judged to occur during the CSE, it was at least four times more likely to occur on VFSE. When thin liquids were judged to be aspirated on, CSE participants were between 6.8 and 9.5 times more likely to aspirate on VFSE. Not all trial swallows measures were rated reliably, but all measures of aspiration from trial swallows were made with good intra- and interjudge reliability.

Two measures ("cough/clear/wet voice" and "oropharyngeal dysphagia") were more global measures relating to the trial swallows and/or CSE as a whole. They were not

Table 5. Sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and interjudge reliability of trial swallow signs for detecting aspiration in a clinical swallowing examination.

	SENS	SPEC	PPV	NPV	+LR	REL[a]
Delayed oral transit	56	71	38	84	1.9	100 (1.0)
5 ml thin liquid	47	79	44	81	2.2	85 (.57)
10 ml thin liquid	40	83	40	83	2.3	85
3-oz swallow	27	93	50	83	3.9	75 (−.17)
Thick liquid	38	82	40	80	2.0	100
Puree	34	74	26	81	1.3	85 (.44)
Solid	39	85	40	84	2.6	92 (.78)
Penetration	87	49	5	92	1.7	91
5 ml thin liquid	64	74	46	86	2.5	92 (.80)
10 ml thin liquid	77	57	33	90	1.8	69 (.40)
3-oz swallow	83	48	29	92	1.6	92 (.81)
Thick liquid	39	85	46	81	2.6	54 (−.09)
Puree	25	87	33	81	1.9	77 (.53)
Solid	28	80	26	82	1.4	100
Aspiration	54	89	62	86	5.1	92 (.80)
5 ml thin liquid	44	94	69	84	6.8	77 (.47)
10 ml thin liquid	38	96	72	85	9.2	92 (.63)
3-oz swallow	48	95	70	88	9.5	100 (1.0)
Thick liquid	21	98	73	80	8.7	77 (.42)
Puree	9	99	60	80	5.7	100 (1.0)
Solid	14	97	50	83	4.2	92
Spontaneous cough	44	82	45	81	2.4	
Spontaneous clear	54	69	37	81	1.7	
Wet voice	63	64	38	83	1.7	
Cough/clear/wet voice	81	47	35	88	1.5	
Oropharyngeal dysphagia	90	53	37	94	1.9	100 (1.0)

Note. Unless otherwise noted, all table values are percentages.

[a]Interjudge reliability is reported as percentage agreement and kappa (when computable).

Table 6. Sensitivity, specificity, and positive likelihood ratio for the best clinical swallowing examination signs for detecting aspiration.

	SENS	SPEC	PPV	NPV	+LR
Weak jaw bilateral	15	99	86	77	17.4
Fail 3 ounce swallow	48	95	70	88	9.5
Fail 10 ml thin liquid	38	96	72	85	9.2
Fail thick liquid	21	98	73	80	8.7
Fail 5 ml thin liquid	44	94	69	84	6.8
Breathy voice	16	98	67	79	6.6
Pneumonia	9	98	67	76	5.8
Weak jaw unilateral	26	96	67	79	5.8
Fail puree	9	99	60	80	5.7
Wet/gurgly voice	22	96	62	80	5.3
Poor oral hygiene	14	97	60	77	4.4
Fail solid	14	97	50	83	4.2
Dysphonia	54	86	54	86	3.8
Strained voice	30	92	52	81	3.6
Drools	23	94	56	78	3.6
Soft palate structure	24	93	53	79	3.4
Nonoral feeding	49	84	51	83	3.1

Note. Except for likelihood ratios, all table values are percentages. Fail = patient equation: Ln = linear equation −2.5005 + 2.6759 (3 oz aspiration) + 1.4326 (weak jaw unilateral) + 1.0258 (dysphonia).

dependent on a particular swallow. Both of these measures are noteworthy in that sensitivity was much higher than for most other individual measures. In theory, this means that negative signs (absence of these signs) are helpful for ruling out aspiration. LRs for these measures are low, as are PPVs, indicating that a positive rating for these measures is less helpful for ruling in aspiration.

Best CSE measures for detecting aspiration. The study variables with the strongest univariate association to the dependent variable are listed in Table 6. There were 17 measures with an LR of at least 3.0. These were entered into a backward, stepwise multiple logistic regression analysis to determine the best model for predicting aspiration. Forty-three aspiration events occurred. To maintain the stability of any regression model, a ratio of 10 events for each entered variable is recommended. Consequently, four variables were selected. To meet the assumptions of the regression procedure, measures that demonstrated significant collinearity (Pearson's r correlations) were removed from the analysis. When measures were interrelated, we evaluated the frequency of occurrence of the sign, theoretic plausibility, and strength of the association to determine inclusion of the variable. For example, although wet/gurgly voice (see Table 6) produced a higher LR (5.3) than a general rating of dysphonia (3.8), wet/gurgly voice was also more highly correlated with other measures, such as poor oral hygiene and nonoral feeding. The remaining variables were entered into the final regression analysis.

A final model was determined on the basis of a review of the goodness-of-fit statistics, percentage of concordance, and change in deviance over parameters fitted. The final model identified the important predictors of aspiration as global judgments of aspiration from the 3-oz swallow test, indicating the participant aspirated; presence of dysphonia; and jaw weakness. After this, a regression equation was derived from the final model that can be tested against other samples for validation of these results: Ln = −2.5005 + 1.4326 (weak jaw unilateral) + 1.0258 (dysphonia) + 2.6759 (3-oz swallow).

Discussion

The purpose of this investigation was to determine the utility of measures used in a CSE for detecting aspiration on VFSE postischemic stroke. No other neurologic etiologies were allowed to participate in this investigation, and no transference of the data can be assumed. We are able to compare these results with results from a previous investigation conducted with the same measures administered in the same manner, all of which have reliability data for support (positive or negative). We cannot overemphasize that detecting aspiration is not the sole objective of the CSE. In practice, the CSE 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. Neither does our research emphasis mean that we consider the CSE a screening examination. According to Nielsen and Lang (1999), a screening examination is performed on asymptomatic participants in the search for subclinical disease. The CSE has a wider range of purposes. One of the most critical purposes is, one could argue, to determine the likelihood that a potentially dysphagic person is aspirating. In present dysphagia practice, such a likelihood is often the motivation for conducting an instrumented examination. Sparing facilities and third-party payers unnecessary expense, and potentially dysphagic people unnecessary travel and radiation exposure associated with the VFSE, may be considered important. That being said, knowledge regarding the presence or absence of aspiration may not help define swallow physiology; neither will it necessarily help define dietary and treatment options.

As stated in the introduction to this article, previous research has suggested that a number of clinical measures may be useful for detecting aspiration. Almost two decades ago, Horner et al. (1988) reported that in acute stroke patients, the presence of dysphonia was predictive of aspiration. More specifically, of 23 patients who aspirated, 21 (about 90%) exhibited dysphonia; of 22 patients who did not aspirate, 15 (about 70%) were

dysphonic. Horner et al. (1993) reported, as well, that dysphonia was an indicator of aspiration. Five years after that, Daniels et al. (1998) reported dysphonia to be one of the six most important measures for detecting aspiration and, in 2001, McCullough et al. also placed it on a list of important measures. Dysphonia in the current investigation was indicative not only of aspiration (LR = 3.8) but also of delayed onset of swallowing (LR = 3.4). According to these results, other, more specific qualifications of voice, such as wet/gurgly, may be even more indicative of aspiration (LR = 5.3) or delayed swallowing (LR = 7.6). Wet/gurgly was not more indicative than dysphonia of aspiration in McCullough et al.'s (2001) study. One difference that could account for this is the number of silent aspirators in each study. Aspiration was silent in less than 30% of the aspirators in McCullough et al.'s (2001) study, meaning that many of the aspirators coughed and potentially cleared remnants of liquid from the vocal cords. In the current investigation, just over half of all participants who aspirated did so silently. This may have allowed more opportunity to hear material in the airway.

Other measures have produced less replicable results. Pharyngeal gag, a measure that was initially reported to be associated with aspiration (Horner et al., 1988; Horner et al., 1993), has been both reaffirmed (Daniels et al., 1998) and categorically denied (Leder, 1996, 1997; Leder & Espinosa, 2002; McCullough et al., 2001). According to Buchholz and Neumann (1997), the truth lies in between. That is, although the gag reflex is not going to separate aspirators from nonaspirators, the gag reflex appears to be one measure with neurophysiological support and some empirical support for inclusion in an overall examination for dysphagia. It is also a standard component in cranial nerve/oral mechanism exams.

Daniels et al. (1998) provided support for both the presence of dysphonia and an abnormal gag as part of a screening for aspiration. They also reported that four other measures—dysarthria, abnormal volitional cough, cough with swallow, and voice change with swallow—were predictive of aspiration and that the presence of two of the six factors represented a markedly increased risk of aspiration. McCullough et al. (2001) evaluated the same six measures with data derived from a similar sample of stroke patients in a VA medical center. Only three of the measures from Daniels et al. (1998)—dysphonia, dysarthria, and cough with swallow—were significantly associated with the presence of aspiration. McCullough et al.'s (2001) results did, however, indicate that the presence of four of the six measures Daniels et al. (1998) proposed provided improved predictability for aspiration. Nonetheless, the mere observance of coughing during or immediately after the swallow was the best single measure for detecting aspiration and was more precise than any of the other measures com-

bined. That does not help, however, with silent aspiration. Leder and Espinosa (2002) examined the same six clinical factors in comparison with FEES and found little evidence to suggest a relationship. Because Leder and Espinosa used a different gold standard (FEES vs. VFSE), and only one of the studies (McCullough et al., 2001) reported reliability for rating clinical measures, the reasons for variability in results remain unknown. The methods for participant recruitment could affect the results. Whereas Leder and Espinosa examined consecutive referrals, Daniels et al. (1998) and McCullough et al. (2001) examined all ischemic stroke patients admitted. That difference can change sensitivity and specificity and, thus, the LR. Days postonset also could play a role. Leder and Espinosa evaluated patients within 24 hr. Daniels et al. (1998) evaluated patients within 5 days of admission. McCullough et al. (2001) enrolled participants up to 6 weeks poststroke, but 90% were within 7 days, and 95% were within 2 weeks. Again, these differences can affect these types of results. Other factors may be site of lesion and percentage of silent aspirators.

In this investigation, the presence of 17 different clinical measures appears to increase the likelihood of aspiration occurring by at least three times (see Table 6). The measures provide LRs in the decreasing order of value. In theory, measures with higher LRs should be worth noting when assessing patients for dysphagia. The LRs reported in Table 6 could be used in conjunction with a nomogram (see Figure 2, which is from Sackett et al., 1998, p. 127) to help determine the posttest probability that an individual patient is aspirating based on the measures used. For example, considering that 26% of all participants in this investigation aspirated, one could estimate that the pretest probability of a stroke patient aspirating is 26%. Using a straight-edge placed on the pretest probability (26%) and extended across the LR for an individual CSE measure—for example, failure on the 3-oz swallow test (LR = 9.5)—the resulting posttest probability (follow the straight-edge through to the posttest column) that the patient aspirates would be about 82%. Whether this indicates that an instrumental examination should be conducted requires a clinical decision. Additional discussion on the use of the nomogram and clinical decision making can be found in Rosenbek, McCullough, and Wertz (2004).

When one extracts for redundancies, frequency of occurrence, and theoretic plausibility, the best measures for detecting aspiration, as determined by regression, are failure of the 3-oz swallow test (LR = 9.5), unilateral jaw weakness (LR = 5.8), and dysphonia (LR = 3.8). The 3-oz swallow test has been previously reported to be a predictor of aspiration poststroke (DePippo et al., 1992; McCullough et al., 2001). Unilateral jaw weakness is the only one of the three measures with no prior mention of importance in the literature. Dysphonia, as discussed above, has

Figure 2. Nomogram for calculating posttest probabilities from pretest probabilities and likelihood ratios. From the Web site of the Centre for Evidence-Based Medicine: http://www.cebm.net/nomogram.asp. Copyright by the Centre for Evidence-Based Medicine, Oxford, United Kingdom. Used with permission.
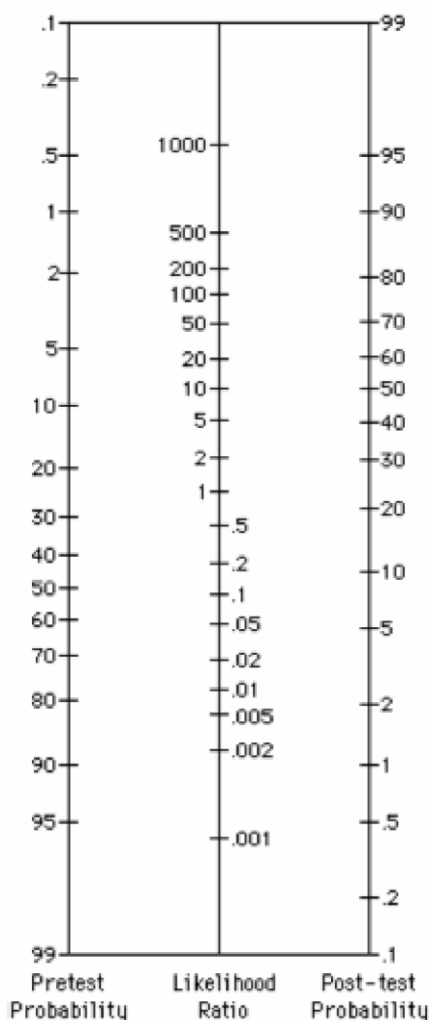
received consistent support for inclusion. Problems with all of these measures should, however, be noted. The 3-oz swallow text cannot (or should not) be administered to everyone. Only individuals who appeared to not aspirate or aspirate minimally during the rest of the CSE were allowed to do the 3-oz swallow test. Thus, only the mildest, or perhaps the most silent, dysphagics took this test. It may be a wonderful test for picking out aspirators from that group, but we argue that other tests should be used before having someone attempt to drink that much water at one time. Similarly, jaw weakness was reported as occurring in only 15 of the participants. Of those, 10 aspirated. Again, even though the presence of jaw weakness should send up a red flag regarding aspiration potential, this would not be useful independently of other measures. Finally, dysphonia occurred in 37 participants. Just over half (20) aspirated. It is significant that dysphonia has continued to be a sign of aspiration and/or dysphagia in numerous studies since the early 1980s (Daniels et al., 1998; Horner et al., 1993; Horner et al., 1988; McCullough et al., 2001). Nonetheless, we are not going to be able to rule out or rule in aspiration on the basis of its presence or absence much better than we are going to rule out "tails" occurring in a coin toss. In fact, no individual measures are going to be able to rule in or rule out aspiration when considered in isolation.

Looking outside our final regression equation may prove more useful. Perhaps the most significant finding in this investigation is that six of the six judgments of aspiration made at bedside (see Table 5) increased the likelihood that aspiration actually occurred on VFSE. LRs ranged from 5.1 to 9.5. This is consistent with previous results (McCullough et al., 2001) in that a clinician's estimation of aspiration is indicative of aspiration on VFSE. This judgment was best following observation of a 10 ml thin liquid and a 3-oz (89 ml) thin liquid swallow (DePippo et al., 1992) using laryngeal palpation with the four-finger method and pre- and postvoicing and were made reliably (more than one clinician made the same judgment using the same measures). Thus, in this one respect, we have replicated an important finding: A well-trained clinician appears to be able to make a statistically accurate judgment that aspiration has occurred in patients who have suffered an acute stroke. This does not mean that a well-trained clinician can detect and rule out aspiration in stroke patients at bedside. It means that, statistically, a well-trained clinician can be right more than wrong in that judgment. Clinically speaking, this may fall short of necessary expectations. Are we missing aspirators at bedside? Yes. Are there negative outcomes associated with the aspirators missed? That question has not been answered. All we know is the majority of these participants (95%) were evaluated within 2 weeks of their stroke, and the CSE measures of note have known reliability among clinicians making the ratings. CSEs and VFSEs were administered in a similar fashion in both this investigation and in McCullough et al.'s (2001) research. We also know that when the study clinicians in these investigations estimated aspiration to occur during bolus swallows, aspiration was between 4.2 and 9.5 times more likely to occur.

It is difficult to determine why this is the case, especially when considering that none of the other specific measures (i.e., cough, wet voice, judgment of delayed oral transit, etc.) were useful. One might conclude that, despite the lack of utility of any single clinical test, the appropriate implementation of a CSE, as used in the current investigation, provides the clinician with sufficient information to make a global judgment regarding the presence of aspiration. Logemann et al. (1999) made a similar statement—that breaking down the CSE into fewer

measures would reduce the overall value of the examination. It is important to remember, however, that sensitivity for all CSE measures was, with few exceptions, relatively low as compared with specificity. Thus, despite a clinician's apparent ability to detect aspiration when it occurs, ruling it out may be more difficult. Two of the more global measures in this investigation—the occurrence of any cough/clear/wet voice and a rating of the presence of oropharyngeal dysphagia—actually appeared to be more useful for ruling aspiration out than ruling it in. Regardless, as a screening tool, ruling out aspiration and ruling it in are equally important. Therefore, the CSE may lack sufficient precision to be a useful screening test. Logemann et al. reported their use of CSE measures for screening patients not only for aspiration but also for delayed swallowing, oral stage problems, and pharyngeal stage problems. Although some of their data support the relationship between clinical signs of swallowing and aspiration, as well as other aspects of dysphagia, the percentage correctly classified was seldom over 70% for any measure. Montgomery and Turkstra (2003) cautioned that "'statistically significant' is not synonymous with 'clinically meaningful'" (p. x). It remains to be determined whether correctly determining aspiration in up to 70% of patients is adequate. Perhaps it is, and the people who are not judged correctly have only a mild dysphagia that will not create negative outcomes. However, this is a supposition in search of empirical support. Age and site of lesion should be considered in this equation as well. Our data are limited in this area but suggest that there are more aspirators poststroke in the 80-and-above age range than in the 79-and-below range. Our data are even more limited regarding site of lesion and aspiration, but bilateral subcortical strokes appeared in conjunction with aspiration in greater percentages (44%) than did unilateral strokes (22%–24%); and the 1 bilateral brain stem stroke patient did aspirate.

In sum, our interpretation of these data is that clinicians cannot rule aspiration in or out using a CSE but that they can use a CSE to gain valuable information regarding aspiration in the context of a CSE that may be used to collect a wider range of information about the patient.

Conclusion

Although a regression equation was derived for the detection of aspiration, a very small number of events contributed to its derivation. Regressions are strongly influenced by prevalence; if 50 more events were added, the resulting equations could be quite different. Furthermore, a lot of associations were observed at the univariate level, extracting measures that would appear to be of some clinical relevance. These regression equations should, therefore, not be considered the most substantive finding of this investigation.

Although none of the measures listed can stand alone as an assessment for aspiration, the presence of many of these measures in the context of a CSE may provide a clinician with valuable information regarding swallowing function. Moreover, the results of this investigation and those of McCullough et al. (2001) indicate that clinicians can provide a bolus to a patient and make a statistically significant judgment regarding aspiration. The reasons for this are not clear, but the four-finger method, pre- and postassessment of voice quality, and other undetermined clinical judgments likely contribute.

Overall, our results imply that some CSE measures serve some purposes. It remains to be determined what the specific purposes are. Currently, we can conclude that the CSE will not detect all aspirators, and it will not rule aspiration out. A 70% correct classification may be inadequate, as may our results that suggest if a clinician judges that a person aspirates during a CSE, it is 10 times more likely that the person will show aspiration on VFSE. Furthermore, even if we detect aspiration successfully with a CSE, we have no idea how much aspiration is occurring and, without physiological information, we are unable to recommend treatments. Additional research is required to answer these questions and to explore the relationship between CSE measures and specific physiological measures of swallowing function, as well as outcome measures. On the basis of the questionable interjudge reliability of many CSE measures, it may also be necessary to provide training to clinicians on the implementation and judgment of all measures used. Until we know how well or how poorly patients continue to do after our CSEs and instrumental evaluations are completed and recommendations have been made, we will not know whether our clinical measures are adequate or whether instrumental measures should always be employed, regardless of the ease or difficulty of obtaining such studies. We do know that clinical practice must be efficient and effective. Costs must be kept down, and outcomes must be kept up. Ultimately, the best way to meet both requirements may be to incorporate the best measures from both instrumental and noninstrumental exams, for example, a clinical FEES examination or a clinical VFSE.

References

Bravo, G., & Hebert, R. (1997). Age- and education-specific reference values for the Mini-Mental and modified Mini-Mental State Examinations derived from a non-demented

elderly population. *International Journal of Geriatric Psychiatry, 12,* 1008–1018.

Buchholz, D., & Neumann, S. (1997). Comments on selected dysphagia literature. *Dysphagia, 12,* 100–105.

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

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

Cogen, R., & Weinryb, J. (1989). Aspiration pneumonia in nursing home patients fed via gastrostomy tubes. *American Journal of Gastroenterology, 84,* 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., Gillespie, M. B., Jones, B., Massey, B. T., Perlman, A. L., & Sonies, B. C. (2003). *Instrumental evaluation of dysphagia.* Paper presented at the annual convention of the Dysphagia Research Society, San Francisco.

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*(4), 17–24.

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.

Elmstahl, S., Bulow, M., Ekberg, O., Petersson, M., & Tegner, H. (1999). Treatment of dysphagia improves nutritional conditions in stroke patients. *Dysphagia, 14,* 61–66.

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.

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.

Howell, D. C. (1992). *Statistical methods for psychology* (3rd ed.). Belmont, CA: Duxbury.

Kertesz, A. (1982). *The Western Aphasia Battery.* New York: Grune & Stratton.

Leder, S. B. (1996). Gag reflex and dysphagia. *Head & Neck, 18,* 138–141.

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., & Espinosa, J. F. (2002). Aspiration risk after acute stroke: Comparison of clinical examination and fiberoptic endoscopic evaluation of swallowing. *Dysphagia, 17,* 214–218.

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 disorder* (2nd ed.). Austin, TX: Pro-Ed.

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

Martino, R., Pron, G., & Diamant, N. (2000). Screening for oropharyngeal dysphagia in stroke: Insufficient evidence for guidelines. *Dysphagia, 15,* 19–30.

McCullough, G. M., Wertz, R. T., & Rosenbek, J. C. (2001). Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. *Journal of Communication Disorders, 34,* 55–72.

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. H., Wertz, R. T., Rosenbek, J. C., Mills, R. H., Ross, K. B., & Ashford, J. R. (2000). Inter- and intra-judge reliability of a clinical/bedside examination of swallowing in adults. *Dysphagia, 15,* 58–67.

Montgomery, E. B., & Turkstra, L. S. (2003). Evidence-based practice: Let's be reasonable. *Journal of Medical Speech–Language Pathology, 11*(2), ix–xii.

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

Robbins, J., & Levine, R. L. (1988). Swallowing after unilateral stroke of the cerebral cortex: Preliminary experience. *Dysphagia, 3,* 11–17.

Rosenbek, J. C., McCullough, G. H., & Wertz, R. T. (2004). Is the information about a test important? Applying the methods of evidence-based medicine to the clinical examination of swallowing. *Journal of Communication Disorders, 37,* 437–450.

Sackett, D. L., Richardson, W. S., Rosenberg, W., & Haynes, R. B. (1998). *Evidence-based medicine: How to practice and teach EBM.* New York: Churchill Livingstone.

Wade, D. T., & Collin, C. (1988). The Barthel ADL Index: A standard measure of physical disability? *International Disability Studies, 10,* 64–67.