Is The Information About A Test Important? Applying The Methods Of Evidence-Based Medicine To The Clinical Examination Of Swallowing

By: John C. Rosenbek, **Gary H. McCullough**, Robert T. Wertz

Abstract

A hotly debated topic in oropharyngeal dysphagia is the Clinical Swallowing Examination's (CSE) importance in clinical practice. That debate can profit from the application of evidence-based medicine's (EBM) principles and procedures. These can guide both appropriate data collection and interpretation as will be demonstrated in the present report. The study's purpose from which data for this report are drawn was to determine the relationship among signs elicited by a CSE and aspiration on a subsequent videofluoroscopic swallowing examination (VFSE). Sensitivity, specificity; positive and negative predictive values (NPV); likelihood ratios; and post-test probabilities for a variety of signs in isolation and in combinations are reported. These data, if judiciously selected and interpreted contribute to the clinician's knowledge about whether to follow a CSE with a VFSE and about what to expect if the VFSE is completed. Learning outcomes: (1) Clinicians will learn how to use EBM principles in conjunction with clinical assessments of swallowing to enhance patient care. (2) Clinicians will learn how to identify combinations of patient signs during he CSE to predict VFSE performance.

John C. Rosenbek, **Gary H. McCullough**, Robert T. Wertz (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 pp. 437-450 Version of Record Available @ www.sciencedirect.com [DOI:10.1016/j.jcomdis.2004.04.007]

The clinical examination for swallowing (CSE) is overshadowed by two instrumental examinations: videofluoroscopy and endoscopy. The reasons are multiple, with a major being the assumption that the CSE fails to provide information on the presence or absence of aspiration on an instrumented examination. Another is that information on the abnormal biomechanics that result in signs of dysphagia and which are critical to treatment planning is unavailable from the CSE. It was this study's purpose to address the first of these reasons: the assumption that data from a CSE are only weakly related to aspiration on the videofluoroscopic swallowing examination (VFSE). Doing so required a systematic approach to the two procedures guided by the principles and procedures of evidence-based medicine (EBM) as they have been applied to a variety of other diagnostic tests. Indeed the title of this paper is a paraphrased borrowing from Sackett, Richardson, Rosenberg, and Haynes (1998): Is the evidence about a diagnostic test important? Specifically the question asked in this study is whether results of the CSE are related to aspiration on VFSE. It is to be emphasized that predicting aspiration on VFSE is but one of the possible reasons for completing a CSE, albeit an important one to a majority of clinicians. Therefore, to answer it, selected data from the CSE and VFSE for 60 patients made dysphagic by stroke were analyzed. After Sackett et al. (1998), the analyses were: true and false negatives and positives, sensitivity and specificity, positive and negative predictive values (NPV), positive and negative likelihood ratios, and pre-and post test probabilities.

1. Procedures

1.1. Participants

Sixty acute stroke patients from two Veteran's Administration (VA) hospitals and one university hospital served as participants. Descriptive data are provided in Table 1. Their mean age is 67.8 with a range of 40 – 96. Fifty-five were male. Mean number of days post-onset was 5.98 with a range of 1 – 42. The majority of these participants were within 2 weeks of stroke. Brain imaging, most frequently with computerized tomography (CT) confirmed that 44 had suffered a single stroke and 16 had suffered two or more strokes. The data on locus of lesion and presence or absence of aspiration appear in Table 2.

Table 1
Descriptive data on study participants

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			

Table 2
Lesion localization and presence or absence of aspiration on VFSE

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

1.2. Clinical swallowing examination (CSE)

Each participant received a CSE comprising four sections: history, oral/motor praxis, voice, and trial swallows. All items were developed from a survey of clinicians' preferences and practices for evaluating dysphagia (McCullough, Wertz, Rosenbek, & Dineen, 1999). All test items reported in this paper meet minimum standards of reliability (Kappa significant at $P < 0.05$) and are summarized in Tables 2 – 5. Procedures have been described previously (McCullough, Wertz, & Rosenbek, 2001) and only a brief summary appears here. To allow for calculation of inter- and intra-judge reliability, three experienced judges reviewed the medical chart to retrieve history data and not surprisingly reliability was high. A variety of verbal and nonverbal tasks were used to elicit oral motor/praxis and voice performance and all three judges working independently scored responses.

Table 3
Sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (-LR) of history signs for detecting aspiration

	SENS (%)	SPEC (%)	PPV (%)	NPV (%)	+LR	-LR
1. Patient report	38	79	43	64	1.8	0.78
2. Family report	14	80	44	41	0.64	1.1
3. Pneumonia	32	92	70	71	4.0	0.74
4. Poor nutrition	50	76	53	91	2.1	0.66
5. Feeding tube	36	95	80	72	6.9	0.67
6. Need suction	5	100	100	65	[a]	0.95
7. COPD	23	82	42	65	1.2	0.95

COPD: chronic obstructive pulmonary disease.
[a] +LR could not be calculated.

Table 4
Sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (+LR), and negative likelihood ratio (-LR) of oral motor signs for detecting aspiration

	SENS (%)	SPEC (%)	PPV (%)	NPV (%)	+LR (%)	-LR (%)
1. Tongue						
(a) Strength	50	74	58	72	1.9	0.7
(b) ROM	36	71	55	66	1.3	0.9
2. Lips						
(a) Strength	27	76	60	66	1.2	1.0
(b) ROM	84	41	44	82	1.4	0.4
3. Jaw						
(a) Strength	[a]	[a]	[a]	61	[a]	[a]
(b) ROM (lateral)	38	74	57	67	1.5	0.8
4. Soft palate						
(a) Strength	41	71	67	68	1.4	0.8
(b) Symmetry	50	53	47	67	1.1	1.0
5. Volitional cough						
(a) Strength	70	45	45	68	1.3	0.7
(b) Quality[b]	55	68	59	73	1.7	0.7
6. Pharyngeal gag	91	18	38	86	1.1	0.5
7. Oral apraxia	41	68	58	67	1.3	0.9
8. Dysarthria	77	55	48	81	1.7	0.4
9. Intelligibility	73	58	50	79	1.7	0.6
10. Secretions	50	84	65	74	3.2	0.6
11. Attends	50	53	47	65	1.1	1.0

ROM: range of motion.

[a] Jaw strength measures could not be calculated due to low numbers of participants with decreased jaw strength.

[b] Wet vs. dry.

Table 5
Sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (þLR), and negative likelihood ratio (-LR) of voice signs for detecting aspiration

	SENS (%)	SPEC (%)	PPV (%)	NPV (%)	þLR	-LR
Speech						
1. Dysphonia	91	22	40	80	1.2	0.4
2. Breathy	82	30	41	75	1.2	0.6
3. Harsh	36	84	40	73	2.2	0.8
4. Wet	50	78	64	73	2.3	0.6
5. Resonance	46	81	57	71	0.4	0.7
Imbalance (sustained AH)						
1. Dysphonia	100	27	43	100	1.4	0.0
2. Breathy	81	37	44	80	1.3	0.5
3. Wet	50	84	67	76	3.1	0.6

Swallowing testing involved the same number and type of boluses used in the videofluoroscopic swallowing examination. The three judges again independently scored performance. Their goal was to listen and look for the signs of aspiration and register a judgment about its presence or absence.

1.3. Videofluoroscopic swallowing examination (VFSE)

One of the authors, G.M., completed all VFSEs within 24 h of the CSE. The protocol has previously been described (McCullough et al., 2001). Each participant received in order: four thin liquids, two 5 cc and two 10 cc (50% barium sulfate powder and 50% water) with a viscosity of 14 cP; two 5 cc thick liquid swallows (mixture of barium sulfate, water and Thick-It) with a viscosity of 187 cP; two 5 cc swallows of applesauce and barium sulfate powder (two tablespoons of barium per four ounces of applesauce); and two cookie swallows (one-fourth of a Lorna Doone coated with applesauce and barium mixture). Only a single judge, G.M., rated the VFSEs within 1 week of their completion and without knowledge of CSE results. Twenty-two patients (37%) were judged to have aspirated (see Table 2).

2. Analysis

Determining if the results of a procedure (e.g., CSE) are useful in detecting a disease or condition (e.g., aspiration on VFSE) begins with the construction of a 2 × 2 table. The 2 × 2 table's purpose is to allow comparison of the test's results with the results on a "gold standard." In the analysis to follow, the data to be reported are the number of participants who aspirated and who did not aspirate on VFSE and the number of those same participants who did and did not have the sign (i.e., wet voice after swallow on the trial swallow portion of the CSE). Looking at Fig. 1 in which these data are used to construct the 2 × 2 table will make the subsequent discussion easier to follow. These are the data that will be used for all subsequent calculations except for the last, when multiple signs (including wet voice after swallow) are combined. Other findings from the CSE could have been used as examples and a variety of tables will summarize specific calculations, for example, sensitivity and specificity and positive and negative predictive values for all reliable data from all four parts of the CSE. What follows, however, are calculations for wet voice and aspiration and no aspiration on VFSE.

2.1. True and false positives and negatives

In clinical practice one hears frequently about results that are true or false positive or true or false negatives. The adjective ***true*** has a good connotation and ***false*** a bad one. In any comparison of a test result with a gold standard, the number of true and false positives and negatives is calculated from a 2 × 2 table. Consider wet voice after swallow and aspiration on VFSE as displayed in Fig. 1. True positives are the number of participants who had wet voice after swallow and who subsequently aspirated on VFSE. That number is recorded in cell "a." Eleven of 60 participants are recorded as true positives. True negatives are the

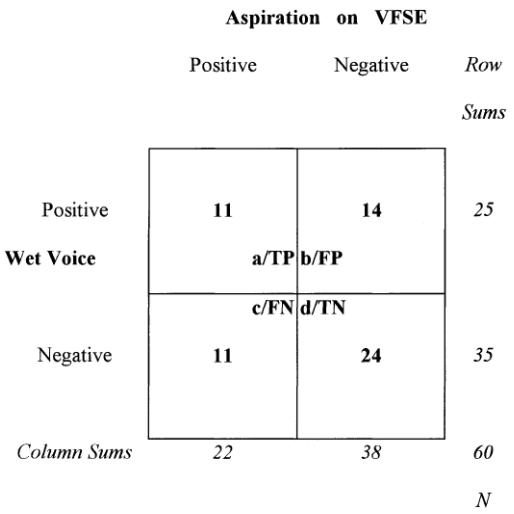

Fig. 1. Signal detection 2 × 2 contingency table demonstrating sensitivity, specificity, positive predictive value, negative predictive value, positive likelihood ratio, and negative likelihood ratio of one clinical/bedside swallowing sign (wet voice after swallow) for detecting aspiration. Sensitivity $=a/(a+c) = 11/(11+11) = 50\%$; specificity $=d/(b+d) = 24/(14+24) = 63\%$; positive predictive value $=a/(a+b) = 44\%$; negative predictive value $=d/(c+d) = 69\%$; positive likelihood ratio = sensitivity/(1 − specificity) = 0.50/(1 − 0.63) = 1.4; negative likelihood ratio = (1 − sensitivity)/specificity ¼ 0.8; CSE ¼ clinical swallowing examination; VFSE ¼ videofluoroscopic swallowing examination; TP = true positive; FP = false positive; FN = false negative; TN ¼ true negative.

number of participants who have normal voice after swallow and who do not aspirate on VFSE. That number recorded in cell ‘d’ is 24. False positives (see cell ‘b’) are those participants (i.e., 14 of the 60 in this study) who have wet voice but do not aspirate on a subsequent VFSE. False negatives (see cell ‘c’) are those participants (i.e., 11 of the 60 in this study) who do not have wet voice after swallow but who subsequently aspirate on VFSE. Fourteen false positives and 11 false negatives from a sample of 60 seem large but taken by themselves these numbers are not as meaningful as they might be. The challenge, therefore, is to use these data to calculate more meaningful values. A number of calculations are available beginning with sensitivity and specificity.

2.2. *Sensitivity and specificity*

Sensitivity and specificity are traditional and widely used calculations to determine the value of a test. The sensitivity of a clinical sign (e.g., wet voice after swallow evidencing aspiration) for detecting a sign on a laboratory measure (e.g., aspiration on the VFSE) is defined as the proportion of patients who have the sign (e.g., aspiration on VFSE), who are also positive for the clinical sign (e.g., wet voice after swallow). Returning to the 4 × 4 table, sensitivity is derived by the following equation: $a/(a+c)$ or the number of true positives divided by the sum of the true positives and false negatives. Therefore, the sensitivity of wet voice after swallow is equal to: 11/(11 +11) or 0.500. This means that out of the people in this sample of stroke patients who actually aspirated, 0.500 (or 50%)

were also positive for ‘‘wet voice after swallow.’’ High values for sensitivity mean that a negative result for the clinical sign should, theoretically, help rule out the diagnosis (Sackett et al., 1998). So if sensitivity for ‘‘wet voice after swallow’’ were 95% rather than 50%, then the lack of a ‘‘wet voice after swallow’’ should help determine that aspiration did not occur. In this example, however, the odds are at chance level.

The specificity of a clinical sign (e.g., wet voice after swallow) is defined as the proportion of patients who do not have the sign on the laboratory measure (e.g., aspiration on VFSE) who also do not have the clinical sign. In our example, the specificity of the sign ‘‘wet voice after swallow’’ for detecting aspiration, as confirmed by VFSE, can be defined as the proportion of patients who do not aspirate on VFSE who are also judged to *not* have ‘‘wet voice after swallow’’ during the clinical exam. Returning to the 2 × 2 table, specificity is derived by the following equation: $d/(b+d)$ or the number of true negatives divided by the total number of false positives and true negatives. Therefore, the specificity of wet voice after swallow is equal to: 24/(14 + 24) or 0.630. This means that 63% of the people in this sample of stroke patients who did not aspirate were also negative for ‘‘wet voice after swallow.’’ High values for specificity mean that positive results for a clinical measure or sign should, theoretically, help rule in the diagnosis.

So, if the specificity for ‘‘wet voice after swallow’’ were 95% rather than 63%, then we could more safely assume that the presence of a wet voice meant that aspiration occurred.

Sensitivity and specificity for the reliable variables from the history portion of the CSE are reported in columns 2 and 3 of Table 3. Similar data for results of oral motor testing are summarized in Table 4. Table 5 contains the data for voice testing, and Table 6 contains the data for trial swallows. One can look for favorites in those data. Doing so reveals interesting findings. Consider gag, thought by many to be an extremely important function to test. As can be seen in Table 4, the sensitivity for gag is 91%. This means aspiration is unlikely in a person with a normal gag. Specificity, however, is only 18%. Thus the presence of an abnormal gag leaves the examiner with little information about aspiration. Gag, therefore, depending on its integrity, may be of limited usefulness. Other variables can be evaluated similarly.

‘‘Sensitivity and specificity describe the proportion of positive and negative test results in a population in whom we ***already know*** who has the disease or not’’ (Go, 1998). This

Table 6
Sensitivity (SENS), specificity (SPEC), positive predictive value (PPV), negative predictive value (NPV), positive likelihood ratio (þLR), and negative likelihood ratio (-LR) of trial swallow signs for detecting aspiration

	SENS (%)	SPEC (%)	PPV (%)	NPV (%)	+LR	-LR
1. Delayed swallow	48	68	52	72	1.5	0.8
2. Spontaneous cough	68	82	68	97	3.7	0.4
3. Wet voice	50	63	45	69	1.4	0.8
4. Estimate P/A	77	63	57	83	2.1	0.4
5. 3 ounce swallow[a]	86	50	39	91	1.7	0.3
6. Overall rating of dysphagia	91	47	43	90	1.7	0.2

P/A: penetration/aspiration.

[a] Wet voice or spontaneous cough/clear within 1 min.

requirement of knowing a priori is considered to be a major draw back to these values, as we do not typically know who has the disease before testing. Otherwise, there would be no reason to test. This point is especially cogent in comparisons of clinical and videofluoroscopic swallowing examinations as the issue is whether or not to follow a clinical examination with an instrumented examination such as the VFSE. Of course with the data presented in this paper, clinicians have the data to decide, depending on the purpose of testing, on the usefulness of a variety of clinical signs. However, other calculations, including negative and positive predictive values (PPV) are also available.

2.3. Positive and negative predictive values

As used in traditional medicine, the positive predictive value of a clinical sign can be defined as the proportion of patients who are positive for that sign who also have a particular disease (Go, 1998). In our example, the PPV of the sign ''wet voice after swallow'' for detecting aspiration, as confirmed by VFSE, can be defined as the proportion of patients who have a ''wet voice after swallow'' during the clinical exam who also aspirate on VFSE. Using the same 2 × 2 contingency table (see Fig. 1) and the data in our four squares representing true and false positives and true and false negatives, PPV is derived by the following equation: $a/(a + b)$. In other words the number of true positives is divided by the total number true and false positives. Therefore, PPV is equal to: 11/(11 + 14) or 44%. This means that 44% of the individuals status-post stroke who test positive for ''wet voice after swallow'' will aspirate. So, rather than reporting the number of people who aspirate and have the clinical sign ''wet voice after swallow,'' we can switch the perspective and report how well a positive clinical sign of ''wet voice after swallow'' ***predicts*** aspiration subsequent to stroke.

As used in traditional medicine, the negative predictive value of a clinical sign can be defined as the proportion of patients who are negative for a clinical sign who do not have the disease (Go, 1998). In our example, the NPV of the sign ''wet voice after swallow'' for detecting aspiration, as confirmed by VFSE, can be defined as the proportion of patients who do not test positive for ''wet voice after swallow'' during the clinical exam and who also do not aspirate. Using the 2 × 2 contingency table (see Fig. 1) and the data in our four squares representing true and false positives and true and false negatives, NPV is derived by the following equation: $d/(c + d)$. In other words, this is the number of true negatives divided by the total number of false and true negatives. Therefore, NPV is equal to: 24/(11 + 24) or 69%. This means that 69% of individuals who are status post stroke and do ***not*** have a ''wet voice after the swallow'' will also not aspirate.

Positive and negative predictive values for all observations and signs are reported in columns 4 and 5 of Table 3 through 6. As can be seen in Table 3, the positive predictive value of needing suctioning is 100%. In a larger sample this value may be reduced, however, a patient requiring suctioning is very likely to aspirate. No other sign has an equal value.

2.4. Likelihood ratios (LR)

Another way of using the same data if the evidence from a test is important, is to calculate positive and negative likelihood ratios. Both are derived from sensitivity and specificity. The

formula for a positive likelihood ratio is +LR ¼ sensitivity/(1 − specificity); the formula for a negative likelihood ratio is -LR ¼ (1 − sensitivity)/specificity.

Rather than being expressed as decimal values or percent probabilities, LRs are expressed in the form of how ''likely'' something is to occur. Thus, instead of telling us that someone has a 44% probability of aspirating based on a clinical sign of ''wet voice after swallow,'' LRs tell us that the likelihood of a ''wet voice after swallow'' is 1.4 times greater for someone who aspirates than someone who does not. +LR ¼ sensitivity/(1 − specificity) or 0.50/(1 − 0.63), which equals 1.4. -LR ¼ (1 − sensitivity)/specificity or (1 − 0.50)/0.63, which equals 0.8. Thus, according to this example, the likelihood of ***not*** having a ''wet voice after swallow'' is 0.8 times more likely (meaning ***less*** likely) in someone who aspirates than in someone who does ***not*** aspirate. Likelihood ratios for all results on the CSE are reported in columns 5 and 6 of Table 3 through 6.

2.5. *Post-test probabilities*

Taking the time to calculate LRs can be advantageous for a reason beyond their value. For example, a positive likelihood ratio can be used with varying ***pre-test probabilities*** to help establish ***post-test probabilities***. The issue in this paper is the post-test probability that a particular patient will aspirate on a VFSE. When a clinician sees a dysphagic patient, he may have a general idea of how likely the patient is to aspirate based on the patient's diagnosis. The clinician, therefore, can establish a ''pre-test probability'' based on that general idea. For example, using the sensitivity and specificity of CSE signs established for stroke patients in the previous investigation (McCullough et al., 2001), the overall pre-test probability of aspiration in a stroke patient is 37%. All stroke patients, however, are not equally likely to aspirate, therefore knowing the locus of lesion can improve the pre-test probability. If the patient had a brainstem stroke, the literature (Horner, Massey, Riski, Lathrop, & Chase 1988) suggests the pre-test probability for that patient aspirating is closer to 50%, perhaps higher, depending on the area of the brainstem disrupted (Kim, Chung, Lee & Robbins, 2000). Pre-test probabilities can be further refined from other research articles or experience in your clinical setting.

A likelihood ratio, derived from the CSE, may be utilized in conjunction with this pre-test probability to establish a ***post-test probability***. Post-test probabilities may be derived mathematically (Go, 1998; Sackett, 1985; Sackett et al., 1998) or by using a nomogram (see Fig. 2). The mathematical method requires a number of calculations not yet described. The first is a calculation of the prevalence. Prevalence is defined mathematically as $(a + c/(a + b + c + d))$. Prevalence in our study is 37%. This value is then folded into the formula for pretest odds defined mathematically as prevalence/(1 − prevalence) or 0.37/(1 − 0.37) ¼ 0.59. The next calculation is of post-test odds defined as pre-test odds × likelihood ratio or 0.59 × 1.4 = 0.83. And finally post-test probability is equal to post-test odds/(post-test odds + 1). In our example, this yields 83/(0.83 + 1) = 0.45.

Fortunately, there is a simpler way and that is to use a nomogram (See Fig. 2), which uses the pre-test probability and the likelihood ratio plus a straight edge to derive a

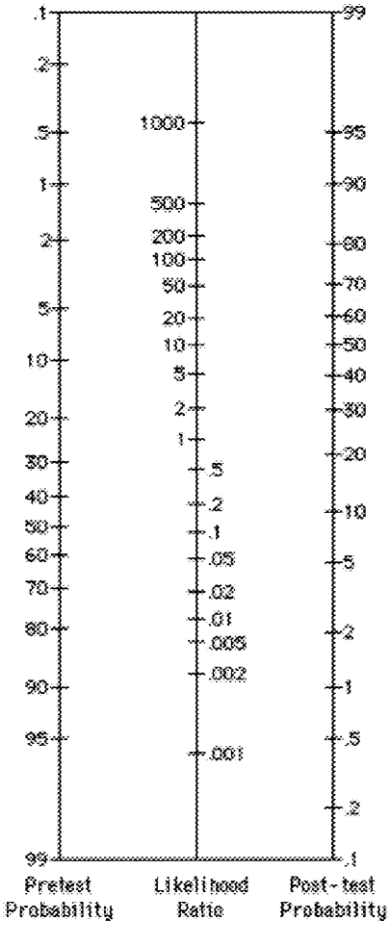

Fig. 2. Nomogram for calculating post-test probabilities from pre-test probabilities and likelihood ratios.

post-test probability. Based on the fact that 37% of the stroke patients in this sample aspirated, we set our pre-test probability at 37%. Place a mark on the left part of the nomogram (Fig. 2, the axis labeled ‘‘pretest probability’’) at 37%. Then, because the +LR for ‘‘wet voice after swallow’’ is 1.4, we place another mark on the center axis of the nomogram (Fig. 2, ‘‘Likelihood Ratio’’) at roughly 1.4. Using a straight edge, we then draw a line from the pretest probability to the likelihood ratio and straight through to ‘‘post-test probability.’’ Using this method, we can determine that the post-test probability for this patient aspirating is about 45%, the same value as that derived mathematically.

Wet voice after swallow on the CSE, despite its clinical appeal, is less than impressive as a sign for predicting aspiration on VFSE. Many of the other findings suffer a similar fate. So what is a clinician to do? One viable alternative is to identify combinations of signs that increase the power of the CSE to predict VFSE performance.

Table 7
Sensitivity (SENS), specificity (SPEC), positive likelihood ratios (þLR) for select, reliable clinical/bedside signs for detecting aspiration

	SENS	SPEC	+LR
History			
1. Pneumonia	0.318	0.921	4.025
2. Poor nutrition	0.500	0.763	2.110
3. Feeding tube	0.364	0.947	6.870
4. Need suction	0.048	1.000	-
Oral motor			
1. Tongue strength	0.500	0.737	0.901
2. Cough strength	0.700	0.450	1.273
3. Secretions	0.500	0.842	3.165
Speech/voice/praxis			
1. Dysphonia	1.000	0.270	1.370
2. Dysarthria	0.773	0.553	1.729
Trial swallows			
Spontaneous cough	0.682	0.816	3.707
Wet voice	0.500	0.632	1.359

2.6. *Combining signs*

Any of the values previously described can be derived for combinations of signs from the CSE. Although one cannot simply add two LRs together to increase likelihood of a positive test result, one can calculate LRs from combinations of measures. To illustrate this methodology, consider the CSE measures with the best +LRs for detecting aspiration (see Table 7). There are four history signs (pneumonia, poor nutrition, feeding tube, need suction), three oral motor signs (tongue strength, cough strength, secretions), two signs extracted from speech/voice/praxis sections (dysphonia, dysarthria), and two trial swallow signs (spontaneous cough, wet voice) which appear better than the other measures in terms of +LR. Using these data we can construct a 2 × 2 contingency table to determine the sensitivity and specificity for various combinations of these signs within each category of the CSE. That number can then be converted into a +LR (+LR ¼ sensitivity/1 − specificity). LRs can then be rank ordered (see Table 8).

If one of four history signs is present, the LR of aspiration is 2.45 times greater than if no history signs were present. As the number of positive signs within a category (i.e., history signs) increases, the LR increases (one of four = 2.45; two of four = 12.23). Therefore, based on these data, if your patient is positive for two of the four history signs, that person would be 12 times more likely to aspirate than someone without two of the four history signs. LRs for three and four history signs present could not be computed because sensitivity reached 1.000, creating an incalculable LR. One can logically assume, however, that having even more of these signs present would increase the likelihood of aspiration or, at least, the need to rule it out instrumentally. One important point to make here is that combined LRs are calculated under different criteria than single measure LRs. For example, Tables 3 and 4 report that the =LR for ‘ ‘pneumonia’ ’ is roughly 4.0. Then, Table 8 reports that the presence

Table 8
Likelihood ratios for combined clinical/bedside signs and nomogram for calculating post-test probability

CSE signs present from four categories	Likelihood ratio
History 2/4	12.23
Signs present from all categories	8.28
History 1/4	8.26
Oral motor 1/3	
Trials swallows 1/2	
History 1/4	6.17
Trial swallows 1/2	
History 1/4	3.97
Oral motor 1/3	
Trial swallows 1/2	2.68
History 1/4	2.45
Oral motor 1/3	2.18
Dysarthria or dysphonia	
Oral motor 1/3	1.96
Dysarthria or dysphonia	1.32

of one of four of these best history signs only provides a =LR of 2.45. How is that possible? It is possible because in the latter example (Table 8) the presence or absence of pneumonia is considered in conjunction with the presence or absence of three other measures. It is a very different analysis. Whether more valuable information can be gained from individual measures or groups of measures remains to be tested; however, since a CSE is comprised of a variety of tasks, one might expect that analyses utilizing more than one variable could prove beneficial. On the other hand, if a clinician assesses swallowing in a patient with a history of pneumonia, the red flag may well be large enough and bold enough to warrant a complete regiment of available assessment tools.

The same methodology for combining signs can be used with oral motor signs, speech praxis/voice, or trial swallows. Moving up from the bottom of Table 8, we observe that if an individual has dysarthria or dysphonia subsequent to stroke, they are only 1.32 times more likely to aspirate, meaning that the presence of either of these two signs may be of limited utility. Other studies have shown stronger relationships between voice disorders and aspiration post stroke (Daniels et al., 1998; Horner et al., 1988). If one of three of the oral motor signs is present, the patient is 1.96 times more likely to aspirate. At this point, signs from different categories can be mixed and matched for analysis. For example, the presence of dysarthria or dysphonia in combination with oral motor signs does indicate an increased risk of aspiration (2.18) over the presence of either of the other signs in isolation (1.32 and 1.96), though not markedly. The presence of one of the four history signs along with one of the four oral motor signs is much more informative (3.97 times more likely to aspirate) as is the combination of one of four history signs with one of the two trial swallow signs (6.17 times more likely to aspirate). If one sign from each of the history, oral motor, and trial swallow signs is present, the likelihood of aspirating post-stroke increases to 8.26. At least one sign present from each of the four categories does not change the results substantively (8.28), which, again, indicates limited utility for measuring dysarthria or dysphonia according to our results.

3. Conclusions

Evidence-based practice depends upon combining clinical experience and insight with the best available data about the features of a particular test for a clinician's specific purpose. Thus there will be no universal answer to the question: Is the evidence about a diagnostic test important? Clinical experience, insight and purposes of testing differ, and nowhere, perhaps, is a universal answer more unlikely than in dysphagia. Training and clinical experience are highly variable and data often go unnoticed. In addition, the reasons for using the CSE differ from clinician to clinician, institution to institution, patient to patient, and even from time to time for the same patient. The data in this paper may be useful to the clinician charged with determining the probability that a stroke patient presenting with a particular history or set of signs on a CSE will aspirate if clinical testing is followed by a VFSE. It is to be recalled that only 60 stroke subjects contributed the data. Data analysis from a replication with a larger sample is under way. Other work with the CSE and VFSE needs to be done. Perhaps foremost is the need to determine if the evidence provided by the CSE can define the biomechanical abnormalities responsible for signs of dysphagia, including, but certainly not limited to, aspiration. In the interim it is critical that the CSE not be relegated to the status of a screening tool. It is far too powerful.

Acknowledgements

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 this paper is to:
   (a) suggest that a clinical examination of swallowing is superior;
   (b) suggest that a clinical examination of swallowing is a waste of time;
   (c) describe how clinicians can use EBM principles in conjunction with clinical assessments of swallowing to enhance patient care;
   (d) discredit the use of 2 × 2 contingency tables in statistical analyses;
   (e) none of the above.
2. High values for sensitivity mean that:
   (a) a negative result for the clinical sign should, theoretically, help rule out the diagnosis;
   (b) a positive result for the clinical signs should help rule in the diagnosis;
   (c) the sign is predictive of aspiration;
   (d) the patient might easily get his feelings hurt;
   (e) none of the above.
3. Likelihood ratios can be derived from:
   (a) 2 × 2 contingency tables;

(b) sensitivity and specificity;
(c) hard, honest work;
(d) a and b;
(e) none of the above.

4. Likelihood ratios:
(a) can be derived for combinations of measures;
(b) can be combined with clinician-based pre-test probabilities to derive a post-test probability;
(c) indicate the ‘‘likelihood’’ that something could occur;
(d) a and b;
(e) all of the above.

5. Evidence-based practice depends upon:
(a) clinical experience;
(b) insight with the best available data about the features of a particular test;
(c) knowledge regarding the principles of EBM;
(d) a and b;
(e) all of the above.

References

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.

Go, A. S. (1998). Refining probability: An introduction to the use of diagnostic tests. In D. J. (Ed.), *Evidence-based medicine: A framework for clinical practice* (p. 21). Stamford, CT: Appleton & Lange.

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.

McCullough, G. H., 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, 33*, 1 – 18.

McCullough, G. H., Wertz, R. T., Rosenbek, J. C., & Dinneen, C. (1999). Clinicians’ preferences and practices in conducting clinical/bedside and videofluoroscopic swallowing examinations in an adult, neurogenic population. *American Journal of Speech Language Pathology, 8*, 149 – 163.

Sackett, D. L. (1985). *Clinical epidemiology: A basic science for clinical medicine*. Boston: Little, Brown, & Co.

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