Screening Accuracy for Aspiration Using Bedside Water Swallow Tests

A Systematic Review and Meta-Analysis

Martin B. Brodsky, PhD; Debra M. Suiter, PhD; Marlís González-Fernández, MD, PhD; Henry J. Michtalik, MD, MPH, MHS; Tobi B. Frymark, MA; Rebecca Venediktov, MS; and Tracy Schooling, MA

BACKGROUND: Hospitalizations for aspiration pneumonia have doubled among older adults. Using a bedside water swallow test (WST) to screen for swallowing-related aspiration can be efficient and cost-effective for preventing additional comorbidities and mortality. We evaluated screening accuracy of bedside WSTs used to identify patients at risk for dysphagia-associated aspiration.

METHODS: Sixteen online databases, Google Scholar, and known content experts through May 2015 were searched. Only prospective studies with patients ≥ 18 years of age given WST screenings validated against nasoendoscopy or videofluoroscopy were included. Data extraction used dual masked extraction and quality assessment following Meta-analysis of Observational Studies in Epidemiology guidelines.

RESULTS: Airway response (eg, coughing/choking) with or without voice changes (eg, wet/gurgly voice quality) was used to identify aspiration during three different bedside WSTs. Pooled estimates for single sip volumes (1-5 mL) were 71% sensitive (95% CI, 63%-78%) and 90% specific (95% CI, 86%-93%). Consecutive sips of 90 to 100 mL trials were 91% sensitive (95% CI, 89%-93%) and 53% specific (95% CI, 51%-55%). Trials of progressively increasing volumes of water were 86% sensitive (95% CI, 76%-93%) and 65% specific (95% CI, 57%-73%). Airway response with voice change improved overall accuracy in identifying aspiration.

CONCLUSIONS: Currently used bedside WSTs offer sufficient, although not ideal, utility in screening for aspiration. Consecutive sips with large volumes in patients who did not present with overt airway responses or voice changes appropriately ruled out risk of aspiration. Small volumes with single sips appropriately ruled in aspiration when clinical signs were present. Combining these bedside approaches may offer improved screening accuracy, but further research is warranted.

CHEST 2016; 150(1):148-163

KEY WORDS: aspiration; dysphagia; pneumonia; review; screening

ABBREVIATIONS: HSROC = hierarchical summary receiver operating characteristic; LR = likelihood ratio; WST = water swallow test

AFFILIATIONS: From the Department of Physical Medicine and Rehabilitation (Drs Brodsky and González-Fernández), the Outcomes After Critical Illness and Surgery Research Group (Dr Brodsky), the Department of Medicine, Division of General Internal Medicine, Hospitalist Program (Dr Michtalik), and the Armstrong Institute for Patient Safety and Quality (Dr Michtalik), Johns Hopkins University, Baltimore, MD; the College of Health Sciences, Division of Communication Sciences and Disorders, University of Kentucky, Lexington,

KY (Dr Suiter); and the National Center for Evidence-Based Practice in Communication Disorders, American Speech-Language-Hearing Association, Rockville, MD (Mss Frymark, Venediktov, and Schooling).

FUNDING/SUPPORT: This study was funded by the National Institutes of Health [Grants 5K23DC013569 (M. B. B.) and 5K23DC011056 (M. G.-F.)]; the Agency for Healthcare Research and Quality [Grant 1K08HS022331 (H. J. M.)]; and the American Speech-Language-Hearing Association's National Center for Evidence-Based Practice in Communication Disorders.

Dysphagia (difficulty swallowing) is a serious condition that affects approximately 9 million adults.¹ Although dysphagia (International Classification of Diseases, 9th Revision code 787.2x) is typically associated with neurologic disorders and head and neck cancers, it is also present in approximately 30% of general hospital admissions² and in approximately 50% of patients at hospital discharge with four of the top five common general medicine hospital diagnoses. Moreover, dysphagia is associated with a 40% increased hospital length of stay³ and an overall cost to US hospitals of as much as \$10 billion annually.^{3,4}

A potentially severe complication of dysphagia is pulmonary aspiration, which may lead to additional morbidities (eg, pneumonia) or even death.^{5,6} Aspiration pneumonia, a potentially preventable hospital-acquired condition,⁷ has estimated incremental costs of \$19,000 to \$25,000 per occurrence, even larger costs when associated with feeding tube placements.^{8,9} Patients with dysphagia are approximately four times more likely to be readmitted within 30 days.¹⁰ Among elderly Medicare patients,

hospitalizations for aspiration pneumonia have doubled,¹¹ with approximately 10% of patients receiving feeding tubes⁹ and 24% associated with death.^{9,11} There is a > 13-fold increased risk of mortality during hospitalization in rehabilitation patients.³ These data, combined with the generally accepted belief that dysphagia is underreported,^{3,4,12} suggest that patients on medical (ie, not neurologic or oncology) floors may not be recognized as having dysphagia, leading to increased hospitalization lengths and increased health-care costs.¹³ Screening for aspiration can, therefore, reduce morbidity, mortality, and costs while increasing health-care quality.

Various tests are used by clinicians to screen patients for dysphagia. To date, most research has not examined the accuracy of specific tests, lacked comparison with an a priori reference standard, and/or lacked a systematic evaluation process,¹⁴⁻²³ leaving clinicians without guidance.²⁴ The aim of this review is to evaluate the clinical utility of a single test, the water swallow test (WST), and determine its diagnostic accuracy for identifying patients who are aspirating.

Methods

Data Sources

We searched 16 databases, Google Scholar, and published research from leading authors in dysphagia screening. Included were English-written, peer-reviewed studies from 1950 through May 2015 reporting sufficient data to calculate metrics of screening accuracy. We included only prospective, observational studies and randomized controlled trials with adult (≥ 18 years of age) patients (Table 1).

Study Selection

Our index test was a WST using any volume of thin liquid. Clinical end points for determining aspiration were observation of a behavioral airway response, specifically coughing, choking, or throat clearing, and/or change in voice (ie, wet/gurgly quality, other change in vocal quality) as perceived by the rater and compared with voice quality before the WST. Our reference standards were nasoendoscopy or videofluoroscopy because of their similar sensitivities in identifying aspiration and universal clinical acceptance.²⁵ Exclusion criteria were as follows: (1) use of an alternative index (eg, pulse oximetry), (2) use of an alternate end point (eg, laryngeal elevation, desaturation), (3) outcome of oral or esophageal dysphagia, or (4) data from unpublished literature.

Data Extraction/Quality Assessment

Two authors (T. S. and R. V.), masked, identified studies for inclusion from the searched collection of abstracts and full-text

articles, and evaluated each study for methodologic rigor.²⁶⁻²⁸ One author (T. B. F.) extracted study data and reported or calculated diagnostic accuracy metrics with results confirmed by an additional author (R. V.). When needed, authors were contacted to provide missing information. All disagreements were resolved by consensus.

Data Analysis

We calculated sensitivity, specificity, and positive and negative likelihood ratios (LRs) from 2×2 tables and generated forest plots to determine pooled accuracies²⁹ using Meta-DiSc version 1.4.³⁰ A continuity correction of 0.5 was used to replace zero cells.^{30,31} Weighting studies for the number of subjects analyzed and random effects modeling were used to adjust for variability across studies.³²

Heterogeneity (measured by I^2 values) was reduced by stratifying WSTs by presentation method (ie, single sips, consecutive sips, progressive volume challenges) and by the patient's clinical response. To avoid bias of repeated analyses of the same patient,³³ we randomly selected a single outcome when studies reported multiple outcomes using the same index test.³⁴⁻³⁷ Causes for heterogeneity were investigated using meta-regression via the MIDAS program³⁸ in Stata version 12.1 (StataCorp), with an a priori significance of $P < .05$. Hierarchical summary receiver operating characteristic (HSROC) curves with both 95% CIs and 95% prediction contours were also created. Publication bias was tested using Deeks test^{39,40} with an a priori significance of $P < .10$.

Results

Search Results

The literature search was completed on May 19, 2015. Twenty-two studies met criteria for review (Fig 1).

CORRESPONDENCE TO: Martin B. Brodsky, PhD, Department of Physical Medicine and Rehabilitation, Johns Hopkins University School of Medicine, 600 N Wolfe St – Phipps 199, Baltimore, MD 21287; e-mail: brodsky@jhmi.edu

Copyright © 2016 American College of Chest Physicians. Published by Elsevier Inc. All rights reserved.

DOI: <http://dx.doi.org/10.1016/j.chest.2016.03.059>

TABLE 1] Literature Search Strategy

Concept	Term(s) Searched/Databases
General search strategy	(diagnostic accuracy) AND (non-imaging clinical swallow exam) AND (adult) AND (dysphagia OR aspiration OR penetration)
PubMed Search Terms	
Diagnostic accuracy	<p>"predictive value of tests"[MeSH] "sensitivity and specificity"[MeSH] "diagnosis"[MeSH] "diagnosis, differential"[MeSH] "diagnostic errors"[MeSH]</p>
Nonimaging clinical swallow exam	<p>"medical history taking"[MeSH] "symptom assessment"[MeSH] "physical examination"[MeSH] "auscultation"[MeSH] "respiratory sounds"[MeSH] "diagnostic self-evaluation"[MeSH] "questionnaires"[MeSH] "video recording"[MeSH] "fluoroscopy"[MeSH] "mass screening"[MeSH] "clinical medicine"[MeSH]</p>
Dysphagia, aspiration, penetration	<p>"deglutition/diagnosis"[MeSH] "deglutition disorders"[MeSH] "pneumonia, aspiration"[MeSH] "respiratory aspiration"[MeSH] "inhalation"[MeSH] "cough"[MeSH]</p>
Database searched	<p>PubMed (NLM) 1947-2015 CINAHL (EBSCO) 1950-2015 Communication & Mass Media Complete (EBSCO) 1950-2015 Education Research Complete (EBSCO) 1950-2015 Health Source: Nursing/Academic (EBSCO) 1950-2015 Psychology and Behavioral Sciences Collection (EBSCO) 1950-2015 PsycINFO (EBSCO) 1950-2015 ScienceDirect 1975-2015 speechBITE (http://www.speechbite.com/) 1950-2015 LILACS (http://lilacs.bvsalud.org/en/) 1982-2015 ProQuest ComDisDome 1950-2015 ProQuest LLBA 1973-2015 ProQuest Social Services Abstracts 1979-2015 REHABDATA (http://www.naric.com/?q=en/REHABDATA) 1981-2014 Web of Science 1950-2015 ClinicalTrials.gov All results as of 2014</p>

References were managed using EndNote X (Thomson Reuters). Full list of articles identified and reasons for inclusion are available on request. MeSH = Medical Subject Headings.

Reliability between masked reviewers for study identification was substantial ($\kappa = 0.776$).⁴¹

Patient Characteristics

Table 2 highlights characteristics of the 4,617 patients included.^{34-37,42-59} Most patients (3,653 of 4,617; 79%) were recruited by medical provider or recruited for suspected dysphagia,^{35,42,44,50,51,55-58} with 3,000 (65%) from one study.⁵⁵ Primary medical diagnosis of patients was largely neurologic (51%), including cerebrovascular accident (34%),^{34-37,42-44,46-51,54,55,58,59} dementia/Alzheimer

disease (2%),^{44,55} degenerative disorders (eg, Parkinson disease; 2%),^{36,44,55} traumatic brain injury (1%),^{35,36,44} and mixed neurologic/neurosurgical diagnoses (13%).^{55,56} Other diagnoses included head and neck cancer (12%)^{35,45,52,53,55,56} and respiratory diseases (10%).^{44,55} The remaining diagnoses were categorized as mixed/other medical/other not specified/unknown (26%).^{35,44,55-58} Patients' mean age when reported was between 46.7 and 75.2 years; one study⁵⁴ reported a median 79 years of age. Most (70%) patients were men. Fifteen (64%) studies were completed in an

Figure 1 – Study selection flowchart.

acute care hospital,^{34,37,42,43,45-49,51-55} seven (32%) studies were completed in a rehabilitation hospital,^{35,36,50,56-59} and one (4%) study was from both of these settings.⁴⁴

Diagnostic Accuracy

Results were analyzed from three different WSTs in which patients were assessed immediately,⁵⁶ at 1 minute,^{36,43,44,51,55,58,59} and at 5 minutes⁴⁷ postswallow for an airway response and/or voice change. The remaining studies either defined time as during or after the swallow^{34,35,37,42,45,48-50,52,54,57} or without reference to time.^{46,53}

Single Sips: Single sips were subgrouped into small volumes (ie, 1-5 mL) and large volumes (ie, 6-20 mL), both accepting airway response and voice change as indicative of aspiration. Based on pooled values, specificities were greater than sensitivities for both small volume (sensitivity, 71%; 95% CI, 63%-78%; specificity, 90%; 95% CI, 86%-93%; positive LR, 6.4; 95% CI, 3.7-11.2)^{46,49,50,57} and large volume trials (sensitivity, 63%; 95% CI, 53%-72%; specificity, 90%; 95% CI, 84%-94%; positive LR, 3.8; 95% CI, 1.7-8.2) (Tables 3, 4).^{42,47,49} These data suggest that single sips provide supporting evidence for correctly identifying patients who are aspirating; however, negative results may be indicative of false negatives.

Consecutive Sips: Consecutive sips (ie, drinking large volumes of 90-100 mL of water without stopping) were subgrouped into studies that only accepted an airway

response^{36,44,48,56,58} and studies that accepted an airway response or voice change as being clinically indicative of aspiration.^{36,43,49,52,55,59} Among the five studies accepting an airway response only, pooled specificity was greater than sensitivity (sensitivity, 54%; 95% CI, 45%-62%; specificity, 78%; 95% CI, 71%-83%; positive LR, 2.5; 95% CI, 1.6-3.9).^{36,44,48,56,58} Conversely, the six studies that accepted both airway response or voice change had a greater pooled sensitivity (sensitivity, 91%; 95% CI, 89%-93%; specificity, 53%; 95% CI, 51%-55%; positive LR, 2.5; 95% CI, 1.8-3.4) (Tables 3, 4).^{36,43,49,52,55,59} Although the data are mixed using this method of presentation, consecutive sips from large volumes without observations of either an airway response or a voice change provide support for the ability to correctly rule out patients who are not aspirating, better than airway response alone.

Progressive Challenges in Volume: A third type of WST progressively challenges patients with increasing volumes of water until either a clinical sign of aspiration is observed or no additional volumes remain in the protocol. Controlled single sip trials leading to consumption of uncontrolled, consecutive large volumes (range, 2-90 mL) were subgrouped into studies that (1) only accepted an airway response, (2) only accepted a voice change, and (3) accepted an airway response or a voice change as being clinically indicative of aspiration. Among the three studies that only accepted an airway response, pooled specificity was greater than sensitivity (sensitivity, 70%; 95% CI, 60%-79%; specificity, 76%;

95% CI, 66%-84%; positive LR, 2.9; 95% CI, 1.9-4.2).^{34,35,37} The four studies that accepted voice change only were similar (sensitivity, 50%; 95% CI, 39%-61%; specificity, 87%; 95% CI, 78%-92%; positive LR, 2.9;

95% CI, 1.7-4.7).^{34,35,37,53} The three studies that accepted airway response or voice change had similar results to consecutive sip trials with the same clinical indicators, with pooled sensitivity greater than specificity

TABLE 2] Participant Characteristics of Included Studies

Study/Year	No.	Mean Age $\pm$ SD, y	Sex, No. of M (No. of F)	Diagnosis	Setting	Diagnostic Category
Chong et al ⁴² /2003	50	Recent stroke: 73.08 $\pm$ 6.33 No recent stroke: 76.21 $\pm$ 6.86	26 (24)	CVA: n = 50	Acute hospital	CVA
Daniels et al ³⁴ /1997	59	66.0 $\pm$ 11.0	59 (0)	CVA: n = 59	Acute hospital	CVA
DePippo et al ⁴³ /1992	44	71.0 $\pm$ 10.0	NR	CVA: n = 44	Acute hospital	CVA
Garon et al ⁴⁴ /1995	100	75.2 $\pm$ 11.3	52 (48)	CVA: n = 50 Dementia: n = 14 Respiratory: n = 13 TBI: n = 5 PD: n = 4 Other: n = 14	Acute hospital Rehabilitation hospital	Mixed
Hassan and Aboloyoun ³⁵ /2014	74	46.7 (NR)	47 (27)	CVA: n = 26 TBI: n = 17 HNC: n = 16 Other: n = 15	Rehabilitation hospital	Mixed
Hey et al ⁴⁵ /2013	80	61.0 $\pm$ 12.9	58 (22)	HNC: n = 80	Acute hospital	HNC
Kidd et al ⁴⁶ /1993	60	72.0 $\pm$ 9.5	25 (35)	CVA: n = 60	Acute hospital	CVA
Lim et al ⁴⁷ /2001	50	67.5 $\pm$ 11.7	31 (19)	CVA: n = 50	Acute hospital	CVA
Mari et al ³⁶ /1997	93	59.8 $\pm$ 16.0	59 (34)	CVA: n = 28 PD: n = 27 TBI: n = 7 MS: n = 7 ALS: n = 6 MD: n = 5 Other degenerative conditions: n = 13	Rehabilitation hospital	Mixed neuro
McCullough et al ⁴⁸ /2001	60	67.8 $\pm$ 9.9	55 (5)	CVA: n = 60	Acute hospital	CVA
McCullough et al ⁴⁹ /2005	165	65.0 $\pm$ 11.0	163 (2)	CVA: n = 165	Acute hospital	CVA
Momosaki et al ⁵⁰ /2013	110	73.0 $\pm$ 10.0	53 (57)	CVA: n = 110	Rehabilitation hospital	CVA
Nishiwaki et al ⁵¹ /2005	61	70.4 (NR)	40 (21)	CVA: n = 61	Acute hospital	CVA
Patterson et al ⁵² /2011 ^a	126	62.0 (NR)	92 (18)	HNC: n = 126	Acute hospital	HNC
Rosen et al ⁵³ /2001 ^b	26	60.0 (NR)	NR	HNC: n = 26	Acute hospital	HNC
Smithard et al ⁵⁴ /1998	94	Median, 79	47 (47)	CVA: n = 94	Acute hospital	CVA
Somasundaram et al ³⁷ /2014	67	68.0 $\pm$ 13.0	46 (21)	CVA: n = 67	Acute hospital	CVA

(Continued)

TABLE 2] (Continued)

Study/Year	No.	Mean Age $\pm$ SD, y	Sex, No. of M (No. of F)	Diagnosis	Setting	Diagnostic Category
Suiter and Leder ⁵⁵ /2008 ^c	3,000	M = 70.1 (NR) F = 66.8 (NR)	1,669 (1,324)	CVA: n = 468 Medical: n = 492 Respiratory: n = 451 Other neurologic: n = 364 HNC: n = 299 Neurosurgery: n = 232 Cardiothoracic: n = 180 Dementia: n = 86 PD: n = 18 Other: n = 391	Rehabilitation hospital	Mixed
Suiter et al ⁵⁶ /2014	25	70.0 (NR)	25 (0)	Medical: n = 8 Neurologic: n = 7 HNC: n = 5 Surgical: n = 5	Acute hospital	Mixed
Wakasugi et al ⁵⁷ /2008	107	69.2 $\pm$ 13.1	83 (24)	Mixed: n = 107	Rehabilitation hospital	Mixed
Wu et al ⁵⁸ /2004	59	71.7 $\pm$ 11.4	44 (15)	CVA: n = 51 Unknown: n = 8	Rehabilitation hospital	Mixed
Zhou et al ⁵⁹ /2011	107	67.8 $\pm$ 13.3	74 (33)	CVA: n = 107	Rehabilitation hospital	CVA

ALS = amyotrophic lateral sclerosis; CVA = cerebrovascular accident; F = female; HNC = head and neck cancer; M = male; MD = myotonic dystrophy; Mixed = neurologic and non-neurologic diagnoses; Mixed neuro = mixed acquired and degenerative neurologic diagnoses; MS = muscular dystrophy; NR = not reported; PD = Parkinson disease; TBI = traumatic brain injury.

^aStudy authors reported sex and age for 110 participants.

^bStudy authors reported mean age for 27 participants; only 26 were analyzed.

^cStudy authors reported sex data for 2,993 participants and diagnostic category for 2,981 participants.

(sensitivity, 86%; 95% CI, 76%-93%; specificity, 65%; 95% CI, 57%-73%; positive LR, 2.3; 95% CI, 1.8-2.9) (Tables 3, 4).^{45,51,54} Again, these results indicate that combining airway response with voice change is more sensitive than using airway response alone.

Hierarchical Summary Receiver Operating

Characteristics: Each of the different screenings, when collapsed across patient responses, offers varied diagnostic accuracy. HSROC curves plot combined sensitivity and specificity for each study and then plot each study as an aggregate point, with 100% sensitivity and 100% specificity being in the upper left corner of the summary plot. The HSROC curves and their summary operating points have lower sensitivities (range, 60%-77%) than specificities (range, 72%-87%) for single sips and progressive volumes, but not for consecutive sips. Single sips, however, had a larger area under the curve (90% vs 81% for consecutive sips and 79% for progressive volumes), suggesting greater accuracy overall. There is also large within-method variability, supporting an overall greater ability to correctly classify patients who are aspirating without effectively reducing false negatives (Fig 2).

Methodologic Quality

All studies provided an adequate study rationale and identified patients through a one-gate procedure with diagnosis unknown at screening (Table 5). Most also provided clear selection criteria (91%) and consecutive sampling for patient enrollment (91%). Methodologic weaknesses included not reporting (27%) or maintaining an appropriate interval between screening and reference test administration (ie, > 48 hours; 14%), independent administration of the index and reference test (45%), and masking of assessors (45%). There was good agreement between masked reviewers for study quality (Table 6).⁴¹

Heterogeneity and Publication Bias

There was a wide degree of inconsistency among studies. A meta-regression of study characteristics revealed four significant sources of heterogeneity: (1) reduced clarity of selection criteria ($P = .02$), (2) inadequate description of measures and procedures for study replication ($P = .04$), (3) no masking ($P = .01$), and (4) inconsistent inclusion of all patients in all analyses ($P = .01$). There was no significant ($P = .246$) publication bias, with the

TABLE 3] Diagnostic Characteristics of Included Studies

Study/Year	No.	Reference and Index Test Characteristics				2 × 2 Table				Aspiration Prevalence, %	Sensitivity		Specificity		Likelihood Ratios			
		Reference Test	Index Test	Index Total Amount, mL	Clinical End Point Response	TP	FP	FN	TN		%	95% CI	%	95% CI	LR+	95% CI	LR–	95% CI
Single sips																		
Chong et al ⁴² /2003	50	FEES	Single sips 10 mL	≤ 50	Airway or voice	27	6	7	10	68	79	0.62-0.91	63	0.35-0.85	2.1	1.0-4.1	0.3	0.2-0.7
Kidd et al ⁴⁶ /1993	60	VFSS	Single sips 5 mL	≤ 50	Airway or voice	20	5	5	30	42	80	0.59-0.93	86	0.70-0.95	5.6	2.4-12.9	0.2	0.1-0.5
Lim et al ⁴⁷ /2001 ^a	50	FEES	Single sips 10 mL	≤ 50	Airway or voice	22	6	4	18	52	85	0.65-0.96	75	0.53-0.90	3.4	1.7-6.9	0.2	0.2-0.8
McCullough et al ⁴⁹ /2005	165	VFSS	Single sips 5 mL	≤ 10	Airway or voice	19	7	24	115	26	44	0.29-0.60	94	0.89-0.98	7.7	3.5-17.0	0.6	0.5-0.8
			Single sips 10 mL	≤ 20		16	5	27	117		37	0.23-0.53	96	0.91-0.99	9.1	3.5-23.3	0.7	0.5-0.8
Momosaki et al ⁵⁰ /2013	110	FEES	Single sips 4 mL	≤ 12	Airway or voice	42	14	3	51	41	93	0.82-0.99	79	0.67-0.88	4.3	2.7-6.9	0.1	0.1-0.3
Wakasugi et al ⁵⁷ /2008	107	VFSS, FEES	Single sips 3 mL	3	Airway or voice	34	2	15	56	46	69	0.55-0.82	97	0.88-1.00	20.1	5.1-79.5	0.3	0.2-0.5
Consecutive sips																		
DePippo et al ⁴³ /1992	44	VFSS	Consecutive sips	90	Airway or voice	16	11	4	13	45	80	0.56-0.94	54	0.33-0.74	1.7	1.1-2.8	0.4	0.1-1.0
Garon et al ⁴⁴ /1995	100	VFSS	Consecutive sips	90	Airway	19	5	35	41	54	35	0.23-0.49	89	0.76-0.96	3.2	1.3-8.0	0.7	0.6-0.9
Mari et al ³⁶ /1997	93	VFSS	Consecutive sips	90	Airway	21	5	21	46	45	50	0.34-0.66	90	0.79-0.97	5.1	2.1-12.4	0.6	0.4-0.8
					Airway or voice	22	7	21	43		51	0.36-0.67	86	0.73-0.94	3.7	1.7-7.7	0.6	0.4-0.8
McCullough et al ⁴⁸ /2001 ^b	60	VFSS	Consecutive sips	90	Airway	19	19	3	19	37	86	0.65-0.97	50	0.33-0.67	1.7	1.2-2.5	0.3	0.1-0.8
McCullough et al ⁴⁹ /2005	165	VFSS	Consecutive sips	90	Airway or voice	21	6	22	116	26	48	0.33-0.65	95	0.90-0.98	9.9	4.3-23.0	0.5	0.4-0.7

(Continued)

TABLE 3] (Continued)

Study/Year	No.	Reference and Index Test Characteristics				2 × 2 Table				Aspiration Prevalence, %	Sensitivity		Specificity		Likelihood Ratios			
		Reference Test	Index Test	Index Total Amount, mL	Clinical End Point Response	TP	FP	FN	TN		%	95% CI	%	95% CI	LR+	95% CI	LR−	95% CI
Patterson et al ⁵² /2011	126	FEES	Consecutive sips	100	Airway or voice	8	27	2	89	8	80	0.44-0.97	77	0.68-0.84	3.4	2.2-5.4	0.3	0.1-0.9
Suiter and Leder ⁵⁵ /2008	3,000	FEES	Consecutive sips	90	Airway or voice	664	1,185	24	1127	23	97	0.95-0.98	49	0.47-0.51	1.9	1.8-2.0	0.1	0.0-0.1
Suiter et al ⁵⁶ /2014	25	VFSS	Consecutive sips	90	Airway	14	4	0	7	56	100	0.77-1.00	64	0.31-0.89	2.6	1.2-5.4	0.1	0.0-0.8
Wu et al ⁵⁸ /2004 ^c	59	VFSS	Consecutive sips	100	Airway	4	10	7	38	19	36	0.11-0.69	79	0.65-0.90	1.7	0.7-4.5	0.8	0.5-1.3
Zhou et al ⁵⁹ /2011	107	VFSS	Consecutive sips	90	Airway or voice	47	31	7	22	50	87	0.75-0.95	42	0.28-0.56	1.5	1.2-1.9	0.3	0.1-0.7
Progressive challenges in volume																		
Daniels et al ³⁴ /1997 ^a	59	VFSS	Progressive amounts 5, 10, 20 mL	≤ 70	Airway	17	7	9	26	44	65	0.44-0.83	79	0.61-0.91	3.1	1.5-6.3	0.4	0.2-0.8
					Voice	7	4	19	29		27	0.12-0.48	88	0.72-0.97	2.2	0.7-6.8	0.8	0.6-1.1
Hassan and Aboloyoun ³⁵ /2014 ^a	74	FEES	Progressive amounts 5, 20, 50 mL	≤ 75	Airway	31	10	11	23	56	74	0.58-0.86	70	0.51-0.84	2.4	1.4-4.2	0.4	0.2-0.7
					Voice	33	9	8	24		81	0.65-0.91	73	0.54-0.87	3.0	1.7-5.3	0.3	0.1-0.5
Hey et al ⁴⁵ /2013 ^a	80	FEES	Progressive amounts 2, 5, 10, 20 mL	≤ 51	Airway or voice	39	16	0	25	49	100	0.91-1.00	61	0.45-0.76	2.5	1.7-3.7	0	0.0-0.3
Nishiwaki et al ⁵¹ /2005	61	VFSS	Progressive amounts 5, 20 mL	≤ 30	Airway or voice	13	14	5	29	30	72	0.47-0.90	67	0.51-0.81	2.2	1.3-3.7	0.4	0.2-0.9
Rosen et al ⁵³ /2001 ^a	26	VFSS	Progressive amounts 5-118 mL	Most < 90	Voice	3	0	7	16	38	30	0.07-0.65	100	0.79-1.00	10.8	0.6-189.7	0.7	0.5-1.1
Smithard et al ⁵⁴ /1998	94	FEES	Progressive amounts 5, 60 mL	≤ 75	Airway or voice	14	25	6	49	21	70	0.46-0.88	66	0.54-0.77	2.1	1.3-3.2	0.5	0.2-0.9

(Continued)

TABLE 3] (Continued)

Study/Year	No.	Reference and Index Test Characteristics				2 × 2 Table				Sensitivity		Specificity		Likelihood Ratios			
		Reference Test	Index Test	Index Total Amount, mL	Clinical End Point Response	TP	FP	FN	TN	Aspiration Prevalence, %	%	95% CI	%	95% CI	LR+	95% CI	LR−
Somasundaram et al ^{37/2014} ^a	67	VFSS	Progressive amounts 5, 10, 20 mL	≤ 50	Airway	21	5	9	21	54	70	0.51-0.85	81	0.61-0.93	3.6	1.6-8.3	0.4
					Voice	1	1	10	21	33	9	0.00-0.41	96	0.77-1.00	2.0	0.1-29.0	1.0
																0.8-1.2	

airway = coughing, choking, and/or throat clearing during or after swallow; FEES = fiberoptic endoscopic swallow study; FN = false negative; FP = false positive; LR = negative likelihood ratio; LR+ = positive likelihood ratio; TN = true negative; TP = true positive; VFSS = videofluoroscopic swallow study; voice = wet voice, gurgly voice or change in voice quality during or after swallow.

^aAuthors responded to inquiry and provided relevant information.

^bAuthors responded to inquiry, but no relevant information was received.

^cAuthors did not respond to inquiry for additional information.

omission of one influential study with 3,000 patients⁵⁵ from the analysis (Fig 3). Comparing funnel plots, this disproportionately large study strongly influenced the results.

Discussion

The goal of this review is to identify a single screening tool that can be used to identify aspiration—one of the potentially medically and financially injurious results of dysphagia. Despite multidisciplinary interest in screening and > 30 years of research and prompts by the Agency for Healthcare Research and Quality and the Joint Commission for guidance, clinical uncertainty for aspiration risk remains.^{4,60-62} The clinical focus and purpose of the WST is to determine whether a patient is aspirating (ie, safety for oral consumption) and/or requires instrumental assessment. An appropriately sensitive screening will benefit health care by reducing costs with unnecessary referrals for instrumental procedures. Our analyses suggest the WST, when administered as consecutive sips from large volumes, offers the best characteristics to rule out overt aspiration (ie, a readily observable airway response, voice change associated with swallowing). When presented as single sips of small and large volumes, the WST offers the best evidence for the ability to correctly classify patients who are aspirating. These data suggest that single sips of any volume provide supporting evidence for correctly identifying patients who are aspirating; however, negative results may be indicative of false negatives.

Patients silently aspirating continue to remain difficult to identify bedside and are at the greatest medical risk for aspiration pneumonia.^{6,63} We were surprised by the omission of silent aspiration data as an outcome of the reference test related to the WST in the peer-reviewed studies included in this review, especially when this clear limitation of WSTs may result in false negatives leading to aspiration pneumonia. Although silent aspiration risk may be dependent on the volume of water presented to the patient,⁶⁴ more research is needed to validate this theory. Adverse events and outcomes from missing silent aspiration on screenings are beginning to emerge.^{65,66}

The goal of a screening test is to identify all patients at risk for a certain condition/disease (true positives) while reducing the number of patients who falsely test positive. Our review suggests that a single type of WST may not offer the level of discrimination clinicians require, but there are merits to its continued use and investigation.

TABLE 4] Diagnostic Accuracy of Water Swallow Tests to Predict Aspiration

Method of Presentation and Clinical Sign of Aspiration Observed	No. of Patients	% Sensitivity (95% CI)	% Specificity (95% CI)	Positive LR (95% CI)	I^2 , %	Negative LR (95% CI)	I^2 , %
Single sips							
1-5 mL							
Airway response or voice change ^{48,51,52,59}	442	71 (63-78)	90 (86-93)	6.4 (3.7-11.2)	48	0.3 (0.1-0.6)	87
6-20 mL							
Airway response or voice change ^{44,49,51}	265	63 (53-72)	90 (84-94)	3.8 (1.7-8.2)	68	0.4 (0.2-0.8)	78
Consecutive sips							
Airway response ^{38,46,50,58,60}	337	54 (45-62)	78 (71-83)	2.5 (1.6-3.8)	50	0.6 (0.4-0.9)	63
Airway response or voice change ^{38,45,51,54,57,61}	3,535	91 (89-93)	53 (51-55)	2.5 (1.8-3.4)	83	0.3 (0.1-0.9)	97
Progressive challenges in volume							
Airway response ^{36,37,39}	189	70 (60-79)	76 (66-84)	2.9 (1.9-4.2)	0	0.4 (0.3-0.5)	0
Voice change ^{36,37,39,55}	192	50 (39-61)	87 (78-92)	2.9 (1.7-4.7)	0	0.7 (0.4-1.1)	88
Airway response or voice change ^{47,53,56}	235	86 (76-93)	65 (57-73)	2.3 (1.8-2.9)	0	0.3 (0.1-0.9)	70

Airway response = presence of coughing, choking, and/or throat clearing before, during, or after swallowing; LR = likelihood ratio; Voice change = perception of a wet voice, gurgly voice, or change in voice quality associated with swallowing.

First, our stratification of the data suggest that consecutive sips of large volumes offer good sensitivity, whereas single sips of small and larger volumes offer good specificity. Therefore, combining presentations of both single sips and consecutive sips from a large volume in a stepwise process may boost sensitivity and specificity within the same patient screening session and warrants further investigation. Of the 22 studies included, only one study provided data to examine this construct, albeit with limitations.⁴⁹ That is, after testing patients with 5 and 10 mL of water, the researchers allowed clinicians to determine whether the 90-mL volume was safe and appropriate based on other portions of the clinical swallow examination. We are not aware of any published WST study using either single or consecutive sips that has identified medical complications.^{36,43,44,48,52,55,56,58,59} However, several studies either warn to avoid large volume consecutive sips in specific patient populations (eg, patients with head and neck cancer, patients with tracheostomy tube)^{56,67} or excluded these patients from participation.^{48,49}

Second, there were three readily observable, clinical responses used to identify aspiration in the WSTs: (1) airway response, (2) voice change, and (3) either airway response or voice change. Singly, airway response or voice change offered only fair degrees of sensitivity and specificity, depending on the method of presentation; however, combining these observations improved

screening accuracy and is likely the mathematical result of increasing the opportunity for true positives and false positives when considering both airway response and voice change together in the same patient. Similar to previous work, these results suggest that accepting multiple clinical signs associated with aspiration offers greater screening accuracy.^{34,35,49,51,68} Further investigation is needed to determine whether combining pulse oximetry with behavioral response improves screening accuracy.

Consistent with previous reviews,^{18,20,69-71} we found considerable variation in methodologic rigor, specifically in reporting time between administration of the WST and reference test, time-to-test (range, 10 minutes⁴² to 7 days⁵¹), and treatment fidelity (ie, all patients being administered full bolus amounts using the same method).^{42,46,49-54,57} As a result, a within-group comparison could not be made because there were too few studies within each index test. Methodologic limitations such as these add to the growing list of factors for the underreporting of dysphagia^{3,12} and potential attenuation of screening accuracy.⁷²

There was a moderate amount of heterogeneity between studies. Variations in enrollment methods within and between studies led to potential sources of bias. Although not statistically tested because of insufficient numbers of studies of homogenous patient populations

and varying index tests even within patient populations, we also suspect that heterogeneity may be attributable to the mix of patients and patient populations within and between studies, especially considering severity of condition and clinical presentation. Despite these limitations, the I^2 values from our analyses were generally low to moderate, suggesting grossly consistent results within study groups. The highest I^2 values were observed with consecutive sips, possibly as the result of five studies adding variability to an otherwise large and well-controlled single study of 3,000 patients.

There was an absence of strong, unequivocal results (high sensitivity with high specificity). Moreover, excluding one study that enrolled 3,000 patients, the remaining 21 studies enrolled between 25 and 165 patients (median, 67 patients), contributing to overall study variability (Fig 2), but these findings are not entirely atypical. A similar variability in study quality has also been noted in the [ClinicalTrials.gov](https://www.clinicaltrials.gov) database.⁷³

Previous and newly published reviews continue to investigate clinical swallow examinations in general, most not controlling for the many different methods used, and all of them offering similar insights and minimal clinical utility.^{14,17-20,71,74,75} This review focuses on a single method, the WST, and controlled for the heterogeneity in the literature using appropriate statistical analyses. Despite these methods, the variability of our results is similar to previous reviews, strongly suggesting that robust statistical methods cannot correct for limitations in experimental methods. Future research in screening and diagnostics related to dysphagia and aspiration must engage large, methodologically rigorous studies and control for rater bias and time between the screening and reference test administrations. Likewise, homogenous patient samples will allow for better interpretable comparisons between patient populations.

Figure 2 – A-C, Hierarchical SROC curves for water swallow tests: single sips (A), consecutive sips (B), and progressive volumes (C). A, Single sips of 1 to 20 mL volumes. Volumes of 1 to 5 mL are combined with volumes of 6 to 20 mL (a); references listed are Chong et al⁴² (1), Kidd et al⁴⁶ (2), Lim et al⁴⁷ (3), McCullough et al⁴⁹ (4), Momosaki et al⁵⁰ (5), and Wakasugi et al⁵⁷ (6). B, Consecutive sips with 90 to 100 mL volumes; references listed are DePippo et al⁴³ (1), Garon et al⁴⁴ (2), Mari et al³⁶ (3), McCullough et al⁴⁸ (4), McCullough et al⁴⁹ (5), Patterson et al⁵² (6), Suiter and Leder⁵³ (7), Suiter et al⁵⁶ (8), Wu et al⁵⁸ (9), and Zhou et al⁵⁹ (10). C, Single sips to consecutive sips of 2 to 90 mL volumes; references listed are Daniels et al³⁴ (1), Hassan and Abolouroun³⁵ (2), Hey et al⁴⁵ (3), Nishiwaki et al⁵¹ (4), Rosen et al⁵³ (5), Smithard et al⁵⁴ (6), and Somasundaram et al³⁷ (7). AUC = area under the curve; SENS = sensitivity; SPEC = specificity; SROC = summary receiver operating characteristic.

TABLE 5] Methodologic Quality of Included Studies

Study/Year	Adequate Rationale	Selection Criteria Clearly Described	Sampling	Study Replication	Independent Administration	Masking	Subject Selection	Test Administration	Avoidance of Bias	Patient Inclusion	Likelihood Ratios and CIs	Risk of Selective Reporting
Chong et al ⁴² /2003	+	+	+	+	0	+	+	+	+	+	+	+
Daniels et al ³⁴ /1997	+	+	+	+	+	+	+	+	+	+	+	+
DePippo et al ⁴³ /1992	+	+	+	+	0	0	+	0	+	+	+	+
Garon et al ⁴⁴ /1995	+	0	+	+	+	0	+	+	+	+	+	+
Hassan and Aboloyoun ³⁵ /2014	+	+	0	+	0	0	+	0	0	0	+	0
Hey et al ⁴⁵ /2013	+	+	+	+	+	+	+	+	+	+	+	+
Kidd et al ⁴⁶ /1993	+	+	+	+	0	0	+	0	+	+	+	+
Lim et al ⁴⁷ /2001	+	+	+	+	+	+	+	+	+	+	+	+
Mari et al ³⁶ /1997	+	+	+	+	+	+	+	0	+	+	+	+
McCullough et al ⁴⁸ /2001	+	+	+	+	0	+	+	+	+	+	+	+
McCullough et al ⁴⁹ /2005	+	+	+	+	+	+	+	+	+	0	+	+
Momosaki et al ⁵⁰ /2013	+	+	+	+	0	0	+	+	+	+	+	+
Nishiwaki et al ⁵¹ /2005	+	+	+	+	0	0	+	0	+	+	+	+
Patterson et al ⁵² /2011	+	+	+	+	0	0	+	0	0	0	+	+
Rosen et al ⁵³ /2001	+	+	+	0	0	0	+	0	+	0	+	0
Smithard et al ⁵⁴ /1998	+	+	+	0	+	+	+	+	0	+	+	+
Somasundaram et al ³⁷ /2014	+	+	+	+	+	0	+	+	+	+	+	+
Suiter and Leder ⁵⁵ /2008	+	+	+	+	0	0	+	+	+	+	+	+
Suiter et al ⁵⁶ /2014	+	+	+	+	+	+	+	+	+	+	+	+
Wakasugi et al ⁵⁷ /2008	+	0	0	0	0	0	+	+	+	+	+	+
Wu et al ⁵⁸ /2004	+	+	+	+	0	0	+	0	+	+	+	+
Zhou et al ⁵⁹ /2011	+	+	+	+	+	+	+	0	+	+	+	+
Total (%)	22 (100)	20 (91)	20 (91)	19 (86)	10 (45)	10 (45)	22 (100)	13 (59)	19 (86)	18 (82)	22 (100)	20 (91)

+ = present; 0 = not present/not stated.

TABLE 6] Kappa Agreement for Individual Study Quality

Quality Indicator	κ Statistic	κ Agreement ^a
1. Was there an adequate and plausible rationale for the study?	1.000	Very good
2. Were the selection criteria clearly described?	1.000	Very good
3. Was a consecutive or random sample of patients enrolled?	0.699	Good
4. Were measures and procedures described well enough for replication?	0.507	Moderate
5. Were the reference standard and index test administered independently (≤ 48 h)?	0.545	Moderate
6. Was masking used to interpret results?	0.723	Good
7. Were participants identified through a one-gate procedure (diagnosis unknown)?	1.000	Very good
8. Was there an appropriate interval between the index and reference test?	0.818	Very good
9. Did all participants receive the reference test (avoidance of bias)?	0.694	Good
10. Were all participants included in the analysis?	0.492	Moderate
11. Were likelihood ratios and confidence intervals reported or calculable?	0.637	Good
12. Were all data stated in the methods collected and reported (risk of selective bias)?	0.645	Good

^aVery good agreement: > 0.80 ; good agreement: 0.60 – 0.80 ; moderate agreement: 0.40 – 0.60 ; fair agreement: 0.20 – 0.40 , and poor agreement < 0.20 .

Study Limitations

This systematic review has two limitations—inclusion of peer-reviewed, published research only and inclusion of studies written in English only. We recognize that although acceptance of peer-reviewed, published

studies ensures that experts scrutinize study quality, the likelihood of publishing studies with high accuracy and significant findings over studies with lower accuracies and nonstatistically significant results leads to a greater risk of publication bias. Although unlikely, studies

Figure 3 – A and B, Deeks funnel plot asymmetry test for publication bias. Studies are represented by red dots with white numbers, with their respective regression lines. The two plots have different y-axis scales, both maximized for viewing purposes. A, Funnel plot includes all 22 studies in this review and demonstrates statistical significance ($P = .011$). References listed are Chong et al⁴² (1), Daniels et al⁴³ (2), DePippo et al⁴³ (3), Garon et al⁴⁴ (4), Hassan and Aboloyoun³⁵ (5), Hey et al⁴⁵ (6), Kidd et al⁴⁶ (7), Lim et al⁴⁷ (8), Mari et al³⁶ (9), McCullough et al⁴⁸ (10), McCullough et al⁴⁹ (11), Momosaki et al⁵⁰ (12), Nishiwaki et al⁵¹ (13), Patterson et al⁵² (14), Rosen et al⁵³ (15), Smithard et al⁵⁴ (16), Somasundaram et al³⁷ (17), Suiter and Leder⁵⁵ (18), Suiter et al⁵⁶ (19), Wakasugi et al⁵⁷ (20), Wu et al⁵⁸ (21), and Zhou et al⁵⁹ (22). B, Funnel plot omits one study³⁵ with 3,000 patients and is not statistically significant ($P = .246$). References listed are Chong et al⁴² (1), Daniels et al⁴³ (2), DePippo et al⁴³ (3), Garon et al⁴⁴ (4), Hassan and Aboloyoun³⁵ (5), Hey et al⁴⁵ (6), Kidd et al⁴⁶ (7), Lim et al⁴⁷ (8), Mari et al³⁶ (9), McCullough et al⁴⁸ (10), McCullough et al⁴⁹ (11), Momosaki et al⁵⁰ (12), Nishiwaki et al⁵¹ (13), Patterson et al⁵² (14), Rosen et al⁵³ (15), Smithard et al⁵⁴ (16), Somasundaram et al³⁷ (17), Suiter et al⁵⁶ (18), Wakasugi et al⁵⁷ (19), Wu et al⁵⁸ (20), and Zhou et al⁵⁹ (21). ESS = effective sample size.

written in other languages could provide additional understanding and/or may contain results that are principally contrary to the findings presented here.

Conclusions

There is merit to use of the WST to rule out overt aspiration using consecutive sips from large volumes. Individual sips best rules in aspiration when clinical signs of aspiration are present, but at the risk of false negatives. The use of progressive changes in volume does not offer sufficient screening accuracy. Combining

airway response with voice response as clinical signs of aspiration improves the diagnostic accuracy for WSTs. For those screening positive, additional testing will be required to discriminate patients who are aspirating from those who are false positive. Our findings suggest that a stepwise approach to screening, using both consecutive sips from large volumes and single sips of any volume, has potential to improve patient care; however, this has not been empirically tested. Additional research supporting this method for screening is necessary.

Acknowledgments

Author contributions: M. B. B. and T. B. F. had full access to all of the data in the study and take responsibility for the integrity of the data and the accuracy of the data analysis. M. B. B., T. B. F., D. M. S., M. G.-F., R. V., and T. S. contributed substantially to the design and conduct of the study and analysis. M. B. B., T. B. F., D. M. S., M. G.-F., H. J. M., R. V., and T. S. contributed substantially to the interpretation of the data and preparation, review, and approval of the manuscript. R. V. and T. S. additionally were responsible for management of the data.

Financial/nonfinancial disclosures: The authors have reported to *CHEST* the following: D. M. S. has received royalties from Springer Publishing. None declared (M. B. B., M. G.-F., H. J. M., T. B. F., R. V., T. S.).

Role of sponsors: The sponsor had no role in the design of the study, the collection and analysis of the data, or the preparation of the manuscript.

Other contributions: We thank the American Speech-Language-Hearing Association (ASHA) information manager, Beverly Wang, MLS, and former ASHA research assistant, Nicole Franceschini, BA, for their contribution to the preparation of this manuscript. Both assisted with the completion of the systematic review literature search. Additionally, Beverly Wang managed all references pertaining to the review and manuscript preparation.

The American Speech-Language-Hearing Association is a nonprofit national professional, scientific, and credentialing association for audiologists, speech-language pathologists, and speech, language, and hearing scientists.

References

- Bhattacharyya N. The prevalence of dysphagia among adults in the United States. *Otolaryngol Head Neck Surg*. 2014;151(5):765-769.
- Cichero JA, Heaton S, Bassett L. Triaging dysphagia: nurse screening for dysphagia in an acute hospital. *J Clin Nurs*. 2009;18(11):1649-1659.
- Altman KW, Yu GP, Schaefer SD. Consequence of dysphagia in the hospitalized patient: impact on prognosis and hospital resources. *Arch Otolaryngol Head Neck Surg*. 2010;136(8):784-789.
- Altman KW. Dysphagia evaluation and care in the hospital setting: the need for protocolization. *Otolaryngol Head Neck Surg*. 2011;145(6):895-898.
- Marik PE. Aspiration pneumonia and aspiration pneumonia. *N Engl J Med*. 2001;344(9):665-671.
- Schmidt J, Holas M, Halvorson K, Reding M. Videofluoroscopic evidence of aspiration predicts pneumonia and death but not dehydration following stroke. *Dysphagia*. 1994;9(1):7-11.
- Centers for Medicare and Medicaid Services. Hospital inpatient PPS for acute care hospitals and the long-term care hospital PPS & FY 2015 rates. CMS-1607-F. Centers for Medicare & Medicaid Services website. <http://www.cms.gov/Medicare/Medicare-Fee-for-Service-Payment/HospitalAcqCond/HAC-Regulations-and-Notices-Items/CMS-1607-F.html>. Accessed June 14, 2015.
- Kozlow JH, Berenholtz SM, Garrett E, Dorman T, Pronovost PJ. Epidemiology and impact of aspiration pneumonia in patients undergoing surgery in Maryland, 1999-2000. *Crit Care Med*. 2003;31(7):1930-1937.
- Siddique R, Neslusan CA, Crown WH, Crystal-Peters J, Sloan S, Farup C. A national inpatient cost estimate of percutaneous endoscopic gastrostomy (PEG)-associated aspiration pneumonia. *Am J Manag Care*. 2000;6(4):490-496.
- Chu LW, Pei CK. Risk factors for early emergency hospital readmission in elderly medical patients. *Gerontology*. 1999;45(4):220-226.
- Baine WB, Yu W, Summe JP. Epidemiologic trends in the hospitalization of elderly Medicare patients for pneumonia, 1991-1998. *Am J Public Health*. 2001;91(7):1121-1123.
- González-Fernández M, Gardyn M, Wyckoff S, Ky P, Palmer J. Validation of ICD-9 Code 787.2 for identification of individuals with dysphagia from administrative databases. *Dysphagia*. 2009;24(4):398-402.
- Cichero JA, Altman KW. Definition, prevalence and burden of oropharyngeal dysphagia: a serious problem among older adults worldwide and the impact on prognosis and hospital resources. *Nestle Nutr Inst Workshop Ser*. 2012;72:1-11.
- Perry L, Love CP. Screening for dysphagia and aspiration in acute stroke: a systematic review. *Dysphagia*. 2001;16(1):7-18.
- Doggett DL, Turkelson CM, Coates V. Recent developments in diagnosis and intervention for aspiration and dysphagia in stroke and other neuromuscular disorders. *Curr Atheroscler Rep*. 2002;4(4):311-318.
- Ramsey DJ, Smithard DG, Kalra L. Early assessments of dysphagia and aspiration risk in acute stroke patients. *Stroke*. 2003;34(5):1252-1257.
- Hines S, Wallace K, Crowe L, Finlayson K, Chang A, Pattie M. Identification and nursing management of dysphagia in individuals with acute neurological impairment (update). *Int J Evid Based Healthc*. 2011;9(2):148-150.
- Daniels SK, Anderson JA, Willson PC. Valid items for screening dysphagia risk in patients with stroke: a systematic review. *Stroke*. 2012;43(3):892-897.
- Schepp SK, Tirschwell DL, Miller RM, Longstreth WT Jr. Swallowing screens after acute stroke: a systematic review. *Stroke*. 2012;43(3):869-871.
- Martino R, Flowers HL, Shaw SM, Diamant NE. A systematic review of current clinical and instrumental swallowing assessment methods. *Curr Phys Med Rehabil Rep*. 2013;1(4):267-279.
- Speyer R. Oropharyngeal dysphagia: screening and assessment. *Otolaryngol Clin North Am*. 2013;46(6):989-1008.
- Etges CL, Scheeren B, Gomes E, Barbosa LD. Screening tools for dysphagia: a systematic review. *Codas*. 2014;26(5):343-349.

23. Kertscher B, Speyer R, Palmieri M, Plant C. Bedside screening to detect oropharyngeal dysphagia in patients with neurological disorders: an updated systematic review. *Dysphagia*. 2014;29(2):204-212.
24. Dollaghan CA. Evidence-based practice in communication disorders: What do we know, and when do we know it? *J Commun Disord*. 2004;37(5):391-400.
25. Langmore SE. Evaluation of oropharyngeal dysphagia: Which diagnostic tool is superior? *Curr Opin Otolaryngol Head Neck Surg*. 2003;11(6):485-489.
26. Dollaghan CA. Appraising diagnostic evidence. In: Dollaghan CA, ed. *The Handbook for Evidence-Based Practice in Communication Disorders*. Baltimore, MD: Paul H. Brookes; 2007:81-104.
27. Whiting P, Rutjes AW, Reitsma JB, Bossuyt PM, Kleijnen J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic accuracy included in systematic reviews. *BMC Med Res Methodol*. 2003;3:25.
28. Whiting PF, Rutjes AW, Westwood ME, et al. QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Ann Intern Med*. 2011;155(8):529-536.
29. Eusebi P. Diagnostic accuracy measures. *Cerebrovasc Dis*. 2013;36(4):267-272.
30. Zamora J, Abraira V, Muriel A, Khan KS, Coomarasamy A. Meta-DiSc: a software for meta-analysis of test accuracy data. *BMC Med Res Methodol*. 2006;6:31.
31. Sweeting MJ, Sutton AJ, Lambert PC. What to add to nothing? Use and avoidance of continuity corrections in meta-analysis of sparse data. *Stat Med*. 2004;23(9):1351-1375.
32. Schmidt FL, Oh IS, Hayes TL. Fixed-versus random-effects models in meta-analysis: model properties and an empirical comparison of differences in results. *Br J Math Stat Psychol*. 2009;62(Pt 1):97-128.
33. Abba K, Kirkham AJ, Oliario PL, et al. Rapid diagnostic tests for diagnosing uncomplicated non-falciparum or Plasmodium vivax malaria in endemic countries. *Cochrane Database Syst Rev*. 2014;12:CD011431.
34. Daniels SK, McAdam CP, Brailey K, Foundas AL. Clinical assessment of swallowing and prediction of dysphagia severity. *Am J Speech Lang Pathol*. 1997;6(4):17-24.
35. Hassan HE, Aboloyoun AI. The value of bedside tests in dysphagia evaluation. *Egyptian Journal of Ear, Nose, Throat and Allied Sciences*. 2014;15(3):197-203.
36. Mari F, Matei M, Ceravolo MG, Pisani A, Montesi A, Provinciali L. Predictive value of clinical indices in detecting aspiration in patients with neurological disorders. *J Neurol Neurosurg Psychiatry*. 1997;63(4):456-460.
37. Somasundaram S, Henke C, Neumann-Haefelin T, et al. Dysphagia risk assessment in acute left-hemispheric middle cerebral artery stroke. *Cerebrovasc Dis*. 2014;37(3):217-222.
38. Dwamena B. MIDAS: A Program for Meta-Analytical Integration of Diagnostic Accuracy Studies in Stata. Ann Arbor, MI: Division of Nuclear Medicine, Department of Radiology, University of Michigan Medical School; 2007.
39. Deeks JJ, Macaskill P, Irwig L. The performance of tests of publication bias and other sample size effects in systematic reviews of diagnostic test accuracy was assessed. *J Clin Epidemiol*. 2005;58(9):882-893.
40. van Enst WA, Ochodo E, Scholten RJ, Hooft L, Leeflang MM. Investigation of publication bias in meta-analyses of diagnostic test accuracy: a meta-epidemiological study. *BMC Med Res Methodol*. 2014;14:70.
41. Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics*. 1977;33(1):159-174.
42. Chong MS, Lieu PK, Sitoh YY, Meng YY, Leow LP. Bedside clinical methods useful as screening test for aspiration in elderly patients with recent and previous strokes. *Ann Acad Med Singapore*. 2003;32(6):790-794.
43. DePippo KL, Holas MA, Reding MJ. Validation of the 3-oz water swallow test for aspiration following stroke. *Arch Neurol*. 1992;49(12):1259-1261.
44. Garon BR, Engle M, Ormiston C. Reliability of the 3-oz water swallow test utilizing cough reflex as sole indicator of aspiration. *J Neurol Rehabil*. 1995;9(3):139-143.
45. Hey C, Lange BP, Eberle S, et al. Water swallow screening test for patients after surgery for head and neck cancer: early identification of dysphagia, aspiration and limitations of oral intake. *Anticancer Res*. 2013;33(9):4017-4021.
46. Kidd D, Lawson J, Nesbitt R, MacMahon J. Aspiration in acute stroke: a clinical study with videofluoroscopy. *Q J Med*. 1993;86(12):825-829.
47. Lim SH, Lieu PK, Phua SY, et al. Accuracy of bedside clinical methods compared with fiberoptic endoscopic examination of swallowing (FEES) in determining the risk of aspiration in acute stroke patients. *Dysphagia*. 2001;16(1):1-6.
48. McCullough GH, Wertz RT, Rosenbek JC. Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. *J Commun Disord*. 2001;34(1-2):55-72.
49. McCullough GH, Rosenbek JC, Wertz RT, McCoy S, Mann G, McCullough K. Utility of clinical swallowing examination measures for detecting aspiration post-stroke. *J Speech Lang Hear Res*. 2005;48(6):1280-1293.
50. Momosaki R, Abo M, Kakuda W, Kobayashi K. Applicability of the two-step thickened water test in patients with poststroke dysphagia: a novel assessment tool for paste food aspiration. *J Stroke Cerebrovasc Dis*. 2013;22(6):817-821.
51. Nishiwaki K, Tsuji T, Liu M, Hase K, Tanaka N, Fujiwara T. Identification of a simple screening tool for dysphagia in patients with stroke using factor analysis of multiple dysphagia variables. *J Rehabil Med*. 2005;37(4):247-251.
52. Patterson JM, Hildreth A, McColl E, Carding PN, Hamilton D, Wilson JA. The clinical application of the 100mL water swallow test in head and neck cancer. *Oral Oncol*. 2011;47(3):180-184.
53. Rosen A, Rhee TH, Kaufman R. Prediction of aspiration in patients with newly diagnosed untreated advanced head and neck cancer. *Arch Otolaryngol Head Neck Surg*. 2001;127(8):975-979.
54. Smithard DG, O'Neill PA, Park C, et al. Can bedside assessment reliably exclude aspiration following acute stroke? *Age Ageing*. 1998;27(2):99-106.
55. Suiter DM, Leder SB. Clinical utility of the 3-ounce water swallow test. *Dysphagia*. 2008;23(3):244-250.
56. Suiter DM, Sloggy J, Leder SB. Validation of the Yale Swallow Protocol: a prospective double-blinded videofluoroscopic study. *Dysphagia*. 2014;29(2):199-203.
57. Wakasugi Y, Tohara H, Hattori F, et al. Screening test for silent aspiration at the bedside. *Dysphagia*. 2008;23(4):364-370.
58. Wu MC, Chang YC, Wang TG, Lin LC. Evaluating swallowing dysfunction using a 100-ml water swallowing test. *Dysphagia*. 2004;19(1):43-47.
59. Zhou Z, Salle J, Daviet J, Stuit A, Nguyen C. Combined approach in bedside assessment of aspiration risk post stroke: PASS. *Eur J Phys Rehabil Med*. 2011;47(3):441-446.
60. Agency for Health Care Policy and Research. *Diagnosis and Treatment of Swallowing Disorders (dysphagia) in Acute-Care Stroke Patients*. Rockville, MD: Agency for Health Care Policy and Research; 1999. Evidence Report/Technology Assessment No. 8 (Prepared by ECRI Evidence-Based Practice Center under Contract No. 290-97-0020).
61. Lakshminarayan K, Rostambeigi N, Fuller CC, Peacock JM, Tsai AW. Impact of an electronic medical record-based clinical decision support tool for dysphagia screening on care quality. *Stroke*. 2012;43(12):3399-3401.
62. Robbins J, Langmore S, Hind JA, Erlichman M. Dysphagia research in the 21st century and beyond: proceedings from Dysphagia Experts Meeting, August 21, 2001. *J Rehabil Res Dev*. 2002;39(4):543-548.
63. Muz J, Mathog RH, Miller PR, Rosen R, Borrero G. Detection and quantification of laryngotracheopulmonary aspiration with scintigraphy. *Laryngoscope*. 1987;97(10):1180-1185.
64. Leder SB, Suiter DM, Green BG. Silent aspiration risk is volume-dependent. *Dysphagia*. 2011;26(3):304-309.
65. Butler SG, Clark H, Baginski SG, Todd JT, Lintzenich C, Leng X. Computed tomography pulmonary findings in

- healthy older adult aspirators versus nonaspirators. *Laryngoscope*. 2014;124(2):494-497.
66. Takahashi N, Kikutani T, Tamura F, Groher M, Kuboki T. Videoendoscopic assessment of swallowing function to predict the future incidence of pneumonia of the elderly. *J Oral Rehabil*. 2012;39(6):429-437.
 67. Leder SB, Suiter DM, Warner HL, Kaplan LJ. Initiating safe oral feeding in critically ill intensive care and step-down unit patients based on passing a 3-ounce (90 milliliters) water swallow challenge. *J Trauma*. 2011;70(5):1203-1207.
 68. Daniels SK, Mahoney MC, Lyons GD. Persistent dysphagia and dysphonia following cervical spine surgery. *Ear Nose Throat J*. 1998;77(6):470, 473-475.
 69. Bours GJ, Speyer R, Lemmens J, Limburg M, de Wit R. Bedside screening tests vs. videofluoroscopy or fibreoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review. *J Adv Nurs*. 2009;65(3):477-493.
 70. Martino R, Pron G, Diamant N. Screening for oropharyngeal dysphagia in stroke: insufficient evidence for guidelines. *Dysphagia*. 2000;15(1):19-30.
 71. Wilkinson AH, Burns SL, Witham MD. Aspiration in older patients without stroke: a systematic review of bedside diagnostic tests and predictors of pneumonia. *Eur Geriatr Med*. 2012;3(3):145-152.
 72. Leeflang MM. Systematic reviews and meta-analyses of diagnostic test accuracy. *Clin Microbiol Infect*. 2014;20(2):105-113.
 73. Califf RM, Zarin DA, Kramer JM, Sherman RE, Aberle LH, Tasneem A. Characteristics of clinical trials registered in ClinicalTrials.gov, 2007-2010. *JAMA*. 2012;307(17):1838-1847.
 74. Al-Hussaini A, Latif EH, Singh V. 12-minute consultation: an evidence-based approach to the management of dysphagia. *Clin Otolaryngol*. 2013;38(3):237-243.
 75. O'Horo JC, Rogus-Pulia N, Garcia-Arguello L, Robbins J, Safdar N. Bedside diagnosis of dysphagia: a systematic review. *J Hosp Med*. 2015;10(4):256-265.