Exploration of the utility of a brief swallow screening protocol with comparison to concurrent videofluoroscopy

Exploration de l'utilité d'un bref protocole de dépistage des troubles de déglutition en comparaison avec une vidéofluoroscopie simultanée

Catriona M. Steele
Sonja M. Molfenter
Gemma L. Bailey
Rebecca Cliffe Polacco
Ashley A. Waito
Dana C. B. H. Zoratto
Tom Chau

KEY WORDS
SWALLOWING
DYSPHAGIA
DEGLUTITION
SCREENING
VIDEOFLUOROSCOPY
ASPIRATION

Catriona M. Steele, Ph.D., CCC-S-LP, BRS-S, ASHA Fellow[1,2,3,4]
Sonja M. Molfenter, M.H.Sc.[1]
Gemma L. Bailey, M.H.Sc.[1]
Rebecca Cliffe Polacco, M.H.Sc.[1]
Ashley A. Waito, B.A.[1]
Dana C. B.H. Zoratto, M.H.Sc.[1,3]
Tom Chau, Ph.D.[3,4]

1. Toronto Rehabilitation Institute, Toronto, ON, Canada
2. Department of Speech-Language Pathology, University of Toronto, Toronto, ON, Canada
3. Institute of Biomaterials and Biomedical Engineering, University of Toronto, Toronto, ON, Canada
4. Bloorview Research Institute, Holland Bloorview Kids Rehab, Toronto, ON, Canada

Abstract

This study involved a direct blinded comparison between the results of a brief, standardized swallowing screening protocol and videofluoroscopy of exactly the same swallows. Forty adults participated. Each participant completed a brief swallow screening protocol involving tongue lateralization, voluntary cough, a voice task, and 2 swallowing tasks (3 swallows of 5cc thin liquid barium suspension and a cup-drinking task). We collected time-linked radiographic data and a high-definition movie of the participant's face, head and neck. The movie data were rated by 7 blinded clinicians (nurses and speech-language pathologists) for evidence of clinical signs associated with dysphagia. The videofluoroscopy data were rated by a separate panel of blinded speech-language pathologists for evidence of penetration-aspiration and post-swallow pharyngeal residues. Predictive statistics were calculated for the movie rating results, compared to the videofluoroscopy results. The results showed that none of the screening questions met our criteria for adequate predictive power: sensitivity, specificity and negative predictive values > 0.6, a false negative rate < 0.2 and a positive likelihood ratio > 1.0. We conclude that swallow screening decisions based on a series of 3-4 thin liquid swallows do not have good clinical utility for detecting dysphagia or penetration-aspiration. We discuss a number of issues in swallow screening research that may have contributed to the difference in these results compared to other studies.

Abrégé

Cette étude portait sur la comparaison à l'aveugle des résultats d'un bref protocole normalisé de dépistage des troubles de déglutition et d'une vidéofluoroscopie des mêmes déglutitions. Quarante adultes y ont participé. Chaque participant a effectué un bref protocole de dépistage des troubles de déglutition, c'est-à-dire une latéralisation de la langue, une toux volontaire, un exercice de voix et 2 exercices de déglutition (trois gorgées barytées de 5cc et un test du verre d'eau). Nous avons recueilli les données radiographiques en ordre chronologique et une vidéo en haute définition du visage, de la tête et du cou des participants. Les données de la vidéo ont été notées à l'aveugle par sept cliniciens (infirmières et orthophonistes) afin de trouver des signes cliniques liés à la dysphagie. Les données de la vidéofluoroscopie ont été notées à l'aveugle par un panel séparé d'orthophonistes afin de trouver des signes de pénétration et d'aspiration de résidus ainsi que des résidus pharyngés post déglutition. Des statistiques de prédiction ont été calculées pour le classement des résultats de la vidéo afin de les comparer à ceux de la vidéofluoroscopie. Les résultats ont démontré qu'aucune des questions du dépistage ne répondait à nos critères de prévisibilité : les valeurs prédictives de la sensibilité, la spécificité et la négativité > 0,6, un taux de faux négatif < 0,2 et un ratio positif de possibilité > 1,0. Nous avons conclu que les résultats d'un dépistage des troubles de déglutition basés sur une série de 3-4 gorgées de liquide clair ne sont pas utiles cliniquement pour détecter la dysphagie ou les troubles de pénétration-aspiration. Nous discutons d'un certain nombre de questions liées à la recherche en dépistage de trouble de déglutition qui ont peut-être contribué à faire la différence entre ces résultats et ceux des autres études.

The early identification of dysphagia and aspiration risk through swallow screening has been recognized as best practice in many guidelines, particularly those applying to the management of stroke (e.g., Canadian Stroke Network, 2005; Joint Commission for the Accreditation of Healthcare Organizations, 2004; Scottish Intercollegiate Guidelines Network, 2004). According to the World Health Organization definition, "Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic"(CCI Conference on preventive aspects of chronic disease, 1951). A screening does not provide sufficient information to support management decisions for those who fail the test by showing evidence of the clinical sign in question; rather, "Persons with positive or suspicious findings must be referred… for diagnosis and necessary treatment" (CCI Conference on preventive aspects of chronic disease, 1951).

Swallow screenings are supposed to be simple tests that can be administered by a variety of trained healthcare professionals. In the ideal world, a dysphagia screening should provide a quick and accurate indication of a patient's risk of aspiration, and the likelihood that they have dysphagia. Existing guidelines fail to clearly define the content and procedures required for valid and reliable swallow screening. For example, the guideline published by the Joint Commission for the Accreditation of Healthcare Organizations (2004), which was removed from their accreditation standards in 2009, left room for a screening protocol to range from very quick and simple tests (Suiter & Leder, 2008) right up to full clinical bedside swallow examinations (Logemann, Veis & Colangelo, 1999).

One response to the mandate to provide swallow screening has been for speech-language pathologists to design protocols for use within their local facilities. A search on the internet, using www.google.ca and the search term "dysphagia screening tool" leads to more than 29,000 results and numerous examples of such tools (e.g., Grey-Bruce Health Network Dysphagia Screening Tool, 2008; Iredell Dysphagia Screen, 2006; Lothian Dysphagia Screening Test, 2005; Oklahoma Dysphagia Screening Tool, n.d.; St. George Dysphagia Screening Tool, n.d.). Researchers have also responded to this mandate by designing and testing different screening protocols. Many of these protocols overlap in their core elements, as shown in Table 1.

Table 1

Comparison of protocol components in previously published swallow screening tools.

Test	Diagnostic Criteria	Readiness/ Alertness Criteria	Secretion Management/ Drooling	Respiratory Rate	Facial Muscle or Oral Motor Testing	Sensory Testing	Baseline Voice Quality Appraisal	Baseline Cough Appraisal	Water Swallows	Post-Swallow Cough or Voice Observation	Other
Standardized Swallow Assessment (Perry, 2001)	Stroke	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Speech criteria
Massey Bedside Swallowing Screen (Massey & Jedlicka, 2002)	Stroke	Yes	Yes	No	Yes	Yes	No	Yes	Yes	Yes	Speech criteria
VAMC Nursing Admission Dysphagia Screening Tool (Bravata et al., 2009)	Acute Ischemic Stroke	Yes	Yes	No	Yes	No	Yes	Yes	Yes	Yes	Patient or family report of dysphagia; speech criteria
Royal Brisbane and Women's Hospital Dysphagia Screening Tool (Cichero, Heaton & Bassett, 2009)	Stroke	Yes	Yes	Yes	No	No	Yes	Yes	Yes	Yes	Patient or family report of dysphagia; speech criteria
TOR-BSST© (Martino et al., 2009a)	Stroke	Implied	No	No	Yes	No	Yes	No	Yes	Yes	

Volume-Viscosity Screening Test (Clave et al., 2008)	Risk for dysphagia	Implied	Yes	Yes	Yes	No	Yes	No	(Yes)	Yes	Pulse Oximetry (≥ 3% drop); begins with nectar at controlled volumes and proceeds based on tolerance
Daniels Swallow Screen (Daniels et al., 1998)	Stroke	Implied	No	No	No	Yes	Yes	Yes	Yes	Yes	Speech criteria
MGH-SST (Cohen, 2008)	Neuroscience admissions	Yes	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes	
3-oz Water Swallow Test (Suiter & Leder, 2008)	No	Implied	No	No	No	No	No	No	Yes	Yes	Failure to complete drinking of 3-oz without stopping
Gugging Swallow Screen (Trapl et al., 2007)	Stroke	Yes	Yes	No	No	No	Yes	Yes	(Yes)	Yes	Begins with saliva swallow followed by semisolid before proceeding to liquid
ASSiST - Acute Swallow Screen in Stroke and TIA (2009)	Stroke	Yes	Yes	Yes	Yes	No	Yes	Yes	Yes	Yes	Speech criteria

Almost all reported swallow screening protocols involve water swallowing, but the precise volume and method of administering water differs across tests. Some tests involve between 3 and 10 repeated sips of small volumes of water (Cohen, 2008, 2009; Martino et al., 2009a). In some cases, initial sips of water are followed by cup drinking (Cohen , 2008; Cichero, Heaton & Bassett, 2009). Other protocols start with small, controlled volumes and increase volume gradually (Clave et al., 2008). Some tests involve the rapid and continuous drinking of a large volume of water (Suiter & Leder, 2008) while others begin with thickened water and progress to thin liquids only when thicker items are tolerated without evidence of difficulty (Clave et al., 2008). Regardless of the specific procedures for administering water, all of these tests look for similar signs of difficulty on water swallowing tasks: a) difficulty completing the task; b) coughing; c) change in voice quality post swallow (specifically vocal wetness); or d) respiratory difficulty.

The predictive power of a screening test is most easily understood by using a two by two contingency table, plotting the test result against the presence or absence of the target problems, which, in the case of swallowing, are aspiration or dysphagia (Sackett, Straus, Richardson, Rosenberg & Haynes, 2000). This allows the calculation of indices such as sensitivity (the proportion of people with the underlying problem who have a positive test result, i.e., they fail the test), specificity (the proportion of those who do NOT have the underlying problem who have a negative test result, i.e. they pass the test), and negative predictive value (the proportion of those who pass the test who do not have the underlying problem). In order for a test to have clinical utility, it should score well on all 3 of these indices. It is also desirable for a test to have a low false-negative rate, so that true cases of the target underlying problem are not missed. A full discussion of these measures can be found in McCullough et al. (2005). Likelihood ratios (which are not susceptible to sample prevalence bias when this does not match prevalence in the broader population) are another important index to include when exploring the utility of a screening measure (Schoenfeld, 2009). Likelihood ratios compare the proportions of patients with and without the disease who have been given the diagnostic test and divide the true-positive rate by the false-positive rate (i.e., sensitivity/1-specificity). Thus, the likelihood ratio represents the probability that a given diagnostic test result would be expected in a patient who has the underlying target disorder (Scherokman, 1997).

Research on swallow screening has attempted to validate specific approaches and demonstrate the predictive power of screening tests (as a whole) and of particular component items, for detecting penetration-aspiration and dysphagia. Several systematic reviews concur that no single specific screening approach or procedural element has adequate predictive power for detecting dysphagia or aspiration (Bours, Speyer, Lemmens, Limburg & de Wit, 2009; Martino, Pron & Diamant, 2000; Perry & Love, 2001). Three kinds of screening validation studies can be found in the literature: a) comparisons against clinical bedside examinations; b) comparisons against flexible endoscopic examinations of swallowing; and c) comparisons against videofluoroscopy (VF). When critically reading the literature on swallow screening performance, clinicians need to ask several questions:

1. Was there any bias or skew in the sample in which the screening test was studied?
2. Are the items in the protocol (individually and combined) logical, reasonable and valid measures for detecting the underlying problem?
3. How much variation was encountered in screening (or validation procedure) results across raters?
4. Were there any differences in the number and types of swallows that were compared from the screening to the validation procedure, and how might this affect the conclusions?
5. What was the delay between the screening and the validation procedure, and is this a potential concern?
6. Were the individuals who judged the screening blinded to the validation procedure (true status) results, and *vice versa*?

Within the literature, some swallow screening studies have been performed exclusively in stroke patients while others have been performed in heterogeneous populations. Similarly, some studies have recruited all incoming stroke patients while others have used convenience samples of referred individuals. Sample-related considerations impact the denominators in sensitivity and specificity calculations, which represent the true occurrence of an abnormal or normal swallowing status. When all other elements are kept equal, studies of people with heterogeneous etiologies (compared to etiologically focused samples), or of all incoming patients within a focused target group (compared to referred samples), are prone to recording lower estimates of the true incidence of an underlying problem, which is likely to increase sensitivity and lower specificity. Conversely, in studies of more focused etiological groups, or of referred samples, the estimated incidence of the underlying problem is likely to be higher than that seen in the broader population, leading to lower sensitivity and increased specificity.

Methodological issues can also impact the reported sensitivity and specificity of a screening tool. For example, the type and number of items included in a screening protocol can influence the power of the test for detecting an underlying problem. Martino, Streiner, Maki & Diamant (2009b) showed that sensitivity for detecting dysphagia improved as the number of water swallows in a screening protocol increased. The chances of a binary verdict that the "problem exists" being correct increases with the number of opportunities that the patient has to demonstrate that problem. Similarly, a recent endoscopic study found that patients who aspirate silently on smaller volumes of liquid are more likely to demonstrate an overt sign of aspiration if a greater volume or number of swallows is obtained (Leder, Suiter and Green, 2010).

Table 2 summarizes the predictive power, validation methods, sample characteristics and use of blinding from a selection of swallow screening test validation studies reported in the literature. With the exception of the Veterans Affairs Medical Center Nursing Admission Dysphagia Screening Tool (Bravata et al., 2009), reported sensitivities are generally quite high across screening studies. On the other hand, specificities are generally poor, even when blinding has been used. This suggests a general trend towards over-identifying dysphagia through swallow screening. While it may be argued that it is preferable to over-identify, rather than under-identify, a health condition that is associated with negative outcomes and health care costs, it may also be argued that over-identification involves unnecessary health-care expenditures, as well as negative quality of life consequences when interventions like diet texture restriction are unnecessarily or overzealously applied.

OBJECTIVES

The purpose of our study was to conduct a direct blinded comparison between the results of a brief, standardized swallowing screening protocol and VF of exactly the same swallows, thereby removing the contributions of time lag and test circumstances to differences in test results. Our hypothesis was that clinically observed signs of swallowing difficulty (i.e. failure of specific questions on the swallow screening observation form) would be associated with the occurrence of physiologically abnormal pharyngeal phase swallowing on the videofluoroscopy, and that this relationship would demonstrate good clinical utility

Table 2

Comparison of the methodology and results of previous swallow screening tool validation studies.

Test	Validation	Population	Sensitivity %	Specificity %	Negative Predictive Value %	Positive Likelihood Ratio	Blinding?
Standardized Swallow Assessment (Perry, 2001)	Chart-documented evidence of dysphagia	Stroke	97	90	Not reported	9.70	Not reported
Massey Bedside Swallowing Screen (Massey & Jedlicka, 2002)	Chart-documented evidence of dysphagia	Stroke	100	100	Not reported	N/A	Not reported
VAMC Nursing Admission Dysphagia Screening Tool (Bravata et al., 2009)	S-LP evaluation of swallowing	Stroke	29	84	68	1.81	Not reported
Royal Brisbane and Women's Hospital Dysphagia Screening Tool (Cichero, Heaton & Bassett, 2009)	S-LP Clinical Swallow Examination and Chart Review	Stroke	95	97	98	31.6	No
TOR-BSST© (Martino et al., 2009a)	VFSS confirmation of dysphagia using P-A Scale and MASA dysphagia subscore	Acute stroke	96	64	93	2.60	Yes
Volume-Viscosity Screening Test (Clave et al., 2008)	VFSS confirmation of aspiration and other abnormal swallowing parameters	Heterogeneous	100	29	Not reported	1.40	Yes
Daniels Swallow Screen (Daniels et al., 1998)	VFSS confirmed aspiration	Acute stroke	92	66	Not reported	4.46	Yes
MGH-SST (Cohen, 2008)	FEES confirmation of dysphagia and/or penetration-aspiration	Neuroscience admissions	89	61	87	2.28	Yes
3-oz Water Swallow Test (Suiter & Leder, 2008)	FEES immediately beforehand	Heterogeneous	96	46	98	1.80	No
Gugging Swallow Screen (Trapl et al., 2007)	FEES measures of aspiration (P-A scale ≥ 5)	Stroke	100	50-69	100	3.23	Yes

Abbreviations: FEES = S-LP = Speech-Language Pathologist; VFSS = Videofluoroscopic Swallowing Study; MASA = Mann Assessment of Swallowing Abilty; VAMC = Veterans Affairs Medical Center; MGH-SST = Massachusetts General Hospital Swallow Screening Tool; Flexible Endoscopic Examination of Swallowing; P-A = Penetration-Aspiration.

of the swallow screening process. We were interested in evaluating the strength of specific clinical signs for detecting underlying problems, and in comparing the clinical utility of screening judgments made by nurses (RNs) with those made by speech-language pathologists (S-LPs). We expected that these would not differ significantly.

METHODS

Participants

Data were collected from a gender-balanced convenience sample of 40 consenting adults (mean age: 67 years), referred for VF at one of two hospitals. Ethical considerations regarding radiation risk led us to use this convenience sample, rather than recruiting individuals from the broader population. Etiologies were mixed, and included inpatients as well as outpatients referred by community physicians for initial investigation of swallowing complaints. Our questions were not specific to any particular diagnostic group. Therefore, medical diagnostic information was not captured, other than to confirm the absence of a history of tracheostomy, head and neck cancer, and any surgery to the head and neck other than routine tonsillectomy or adenoidectomy. The study received human subjects approval from the institutional review boards of the participating hospitals.

Screening Protocol

We selected 5 tasks for inclusion in a brief swallow screening protocol, based on a review of the swallow screening literature (see Table 1). The protocol most closely resembled part 2 of the Massachusetts General Hospital Swallow Screening Test (MGH-SST; Cohen, 2008, 2009), and included the following steps:

a. a tongue lateralization task;
b. a baseline voluntary cough task;
c. a baseline phonation task;
d. a sequence of three single-sip swallow tasks (5 cc per sip) with thin liquid, with each sip followed by a repetition of the phonation task;
e. a cup-drinking task with thin liquid, again followed by a repetition of the phonation task.

A readiness-for-testing component, such as that detailed in part 1 of the MGH-SST, was not specifically included. Given the research requirement that all participants be able to consent to the study, it can reasonably be assumed that all of our participants would have received a passing score on such a component. The selected phonation task was the utterance "ha-ha-ha-ha", as used in the MGH-SST protocol. We excluded the evaluation of pharyngeal sensation based on evidence that this component does not add useful information to screening outcomes (Martino et al., 2009a). Due to the fact that the collection of concurrent VF data involved radiation exposure, the cup-drinking task required the patient to "take several continuous sips" from a cup. This was not a complete 3 oz liquid swallowing challenge, as used in some screening protocols (Cohen, 2008, 2009; Suiter and Leder, 2008; Cichero, Heaton and Bassett, 2009), but is consistent with usual videofluoroscopy procedures (Martin-Harris et al., 2008; Logemann, 1993).

Data Collection Procedures

The experiment was set up in the videofluoroscopy suite, using a high-definition camcorder, positioned to capture an image of the patient's face, head and neck, and a high quality stage microphone to capture sound. A dual-axis accelerometer (Analog Devices, ADXL 322) was also placed on the patient's neck in midline over the cricoid cartilage. Time-linked data collection from all channels was controlled by LabVIEW software (National Instruments, Toronto, Canada,). The VF was captured through the hospital's fluoroscopy equipment at 30 frames per second. Figure 1 shows an example of time-linked movie and VF images for a participant during a thin liquid swallow. The acoustic (microphone) and accelerometry data have been reported elsewhere (Waito, Bailey, Molfenter, Zoratto & Steele, 2010; Zoratto, Chau & Steele, 2010) and were not included in the specific analyses for this manuscript.

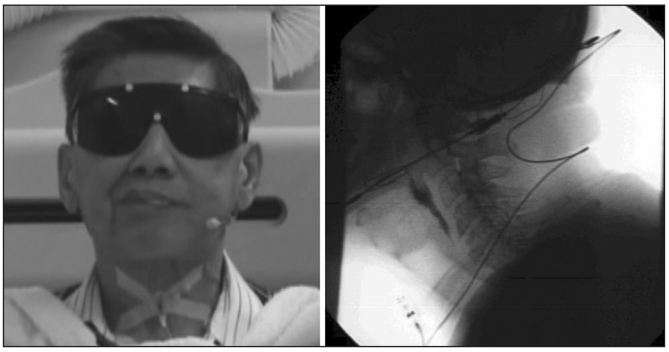

Figure 1. Time-linked screen captures from the movie and videofluoroscopic recording channels, showing a participant swallowing thin liquid barium.

The data collection protocol proceeded as follows:

1. A movie recording was taken while the participant performed the baseline tongue lateralization, volitional cough, and phonation tasks.
2. Time-linked movie and VF recordings were collected during the swallowing of three 5-cc volumes of thin liquid barium suspension (40% w/v), with each swallow followed by repetition of the phonation task. These boluses were administered by teaspoon and a command swallow paradigm was used.

3. A thin liquid barium suspension cup-drinking task (i.e. 3-5 consecutive sips of unrestricted volume) was recorded in the time-linked movie and fluoroscopy channels, again followed by a phonation sample.

The videofluoroscopy then continued for clinical investigative purposes, but these additional swallows were not part of this research study.

Data Processing and Rating

The steps involved in data processing and rating for the screening movies are illustrated in Figure 2. The movies were organized in clips capturing the entire swallow screening sequence, and randomized for rating. Title screens were added to introduce each task in the screening protocol (e.g. "tongue lateralization", "single sip #1"). An algorithm written in MATLAB 2010a (Mathworks, Natick, MA, USA) was used to remove low amplitude background noise from the audio channel of all the movie recordings by filtering out data below 10% of the maximal amplitude found in each signal. This removed contamination from any background conversation and enabled the rater to focus clearly on the voice samples produced by the participant when rating.

The screening movies were rated by 4 S-LPs, in their first year of practice, and 3 registered nurses from an acute care hospital. Using new graduate S-LPs conferred the advantage that none of the raters had any previous knowledge of the participating patients. The acute-care hospital from which the RN judges were drawn was not using a specific swallow screening protocol at the time of the study. We intentionally did not provide extensive didactic training in the identification of the clinical signs of interest because we wanted to determine the utility of the swallow screening tool in the hands of health-care professionals who should, by virtue of their professional knowledge and skills, be able to focus on the specific questions raised, without additional training. Training in the scoring of each sign was conducted with 3 cases not included in the experimental dataset. This allowed us to clarify any questions that arose, and to make sure that all raters were attuned to the signs of interest.

After watching the entire screening movie for a particular participant, the rater was asked to record forced-choice judgments (normal, abnormal, uncertain) regarding the different clinical signs of interest (tongue lateralization; baseline voluntary cough; baseline voice quality; post-swallow spontaneous coughing or change in voice-quality; and overall pass-fail, with fail indicating a need for further swallowing assessment). We asked the raters to give a single rating for each question, across the entire movie sequence for each participant, mirroring the conventional expectation that swallow screenings lead to a single pass-fail result. We took the conservative position of classifying scores of "uncertain" as reflecting the presence of an abnormal clinical sign (i.e., fail) for the purposes of further analysis. This was motivated by the assumption that any question of abnormality should be followed up in the context of usual swallow screening in the clinical setting. The percent agreement across all raters within each professional group and across all 7 raters was calculated for each question for each movie. In cases where there was no clear majority consensus in the screening result for a particular question within a professional group, repeated rating by two of the original raters was conducted.

In addition to providing forced choice answers on the five primary screening questions, raters were also asked to report observations of "any other signs of difficulty" and were given a comments box in which they could elaborate on these observations.

The VF ratings were treated in a similar manner, with the important distinction that the recordings were spliced into clips, randomized and rated at the level of the individual swallowing task (single 5cc sip; cup drinking task). This decision was motivated by the desire to collect an accurate gold-standard answer regarding the presence or absence of dysphagia and aspiration, which would not be biased by knowledge regarding previous swallows

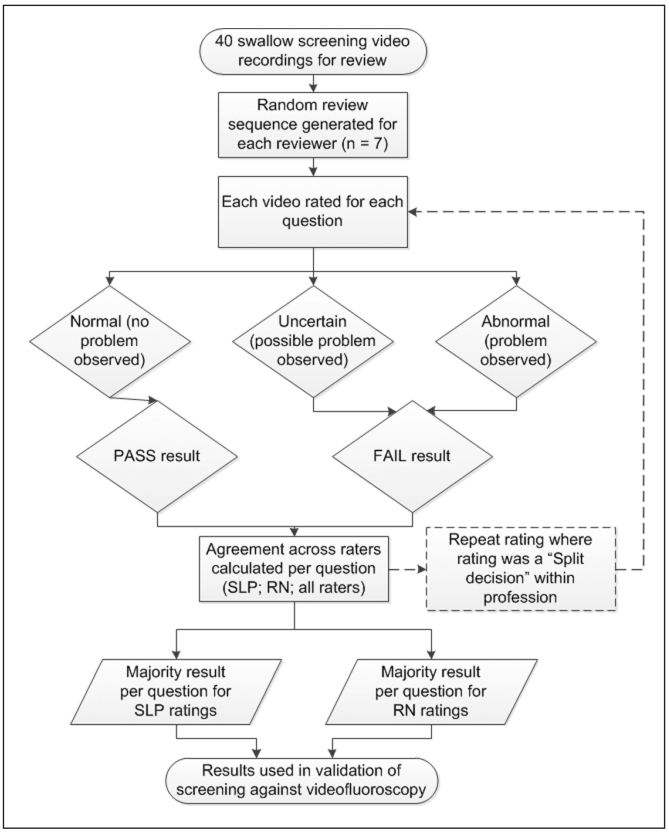

Figure 2. Flow-chart illustrating the data processing and rating steps for determining swallow screening results based on the movies of the swallow screening sequence.

during rating. For 3 participants, video quality concerns (such as shoulder interference or a premature turn-off of the fluoroscopy stream) resulted in the availability of only 1 single sip recording. In a further 9 cases, only 2 single sip clips were of adequate quality to permit rating. Consequently, the VF data set included a total of 106 X 5cc single sip swallows. Cup drinking sequences were available for 36 participants.

The VF rating process is illustrated in Figure 3. A panel of 7 S-LPs (separate from those involved in the movie ratings) completed a 6-hour training session in the rating procedures prior to the study. The individual swallow clips were rated using the 8-point Penetration-Aspiration Scale (Rosenbek, Robbins, Roecker, Coyle & Wood, 1996) and Eisenhuber's ordinal scales capturing residues in the valleculae and pyriform sinuses, with each residue location scored separately (Eisenhuber et al., 2005). The rating assignment (a total of 1275 feature ratings) was randomized across raters, and was organized in batches (10 swallow clips per batch), for which a rater was required to rate a single feature (aspiration or residue). The final data set included 3 ratings per feature for each swallow. These ratings were reviewed for concordance across raters. For 6 clips, consensus was not established in the initial ratings of penetration-aspiration, while 3 and 5 clips, respectively, showed a lack of consensus for vallecular and pyriform sinus residue ratings. These clips were entered into a new rating set, which was reviewed in a live consensus session attended by two of the original raters and a research assistant facilitator. Each clip was played 3 times and the raters recorded scores independently. Scores were then declared. In the case that these new scores differed, the clip was reviewed carefully and discussed until an agreed score was reached.

Once the VF ratings were complete, binary penetration-aspiration and cumulative residue disposition scores were derived for each participant across all 4 swallowing tasks in the screening sequence. In this way, the resolution of the binary screening and VF verdicts was equalized. Any occurrence of penetration-aspiration scale scores ≥ 3 (i.e. material entering the supraglottic space without subsequent ejection) resulted in a disposition classification of "penetration-aspiration present". A cumulative residue score (i.e. vallecular residue score plus pyriform sinus residue score) ≥ 2 (representing either 2 or more occurrences of mild residue or at least one occurrence of moderate residue) resulted in a disposition classification of "residue present". A classification of "dysphagia present" was assigned in the case of either "penetration-aspiration present" and/or "residue present". As such, the threshold for classifying a patient as truly having dysphagia was low, with the exception that a single episode of mild residue was considered insufficient to classify a person as having dysphagia.

ANALYSIS

We calculated the predictive value of each abnormal screening test result (from the screening movie ratings) for dysphagia and penetration-aspiration on VFSS, looking separately at the results for the RN and S-LP raters. Two-by-two contingency tables were prepared for each of the five movie rating results, compared to the videofluoroscopic dispositions of aspiration-present and dysphagia-present. Data were analyzed at the level of the individual patient (across the series of thin liquid swallows and non-swallowing tasks collected during the screening protocol), rather than on a swallow-by-swallow basis. The criterion of considering a particular result to have adequate predictive utility was defined a *priori* as the combination of sensitivity, specificity and negative predictive value scores > 0.6, a false negative rate < 0.2 and a positive likelihood ratio > 1.0.

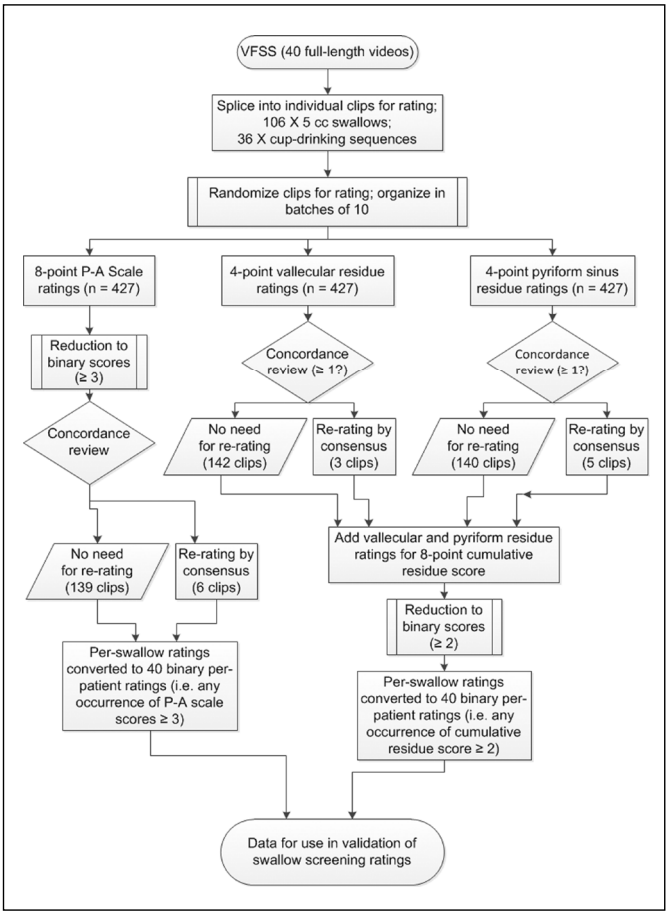

Figure 3. Flow-chart illustrating the data processing and rating steps for determining true aspiration and dysphagia status based on blinded review of the videofluoroscopies.

RESULTS

Descriptive statistics for the frequency (percent occurrence) of abnormal findings in the movie ratings are shown by rater group (S-LPs, RNs, combined) in Table 3. In Table 4, the intra-class correlations for inter-rater agreement in the binary disposition scores for each question are shown by rater group. Table 4 also shows the agreement between professions for the final disposition scores assigned to each participant, considering all raters within each professional group. This agreement was generally strong, as indexed by Cohen's Kappa scores of $\kappa > 0.6$. The kappa score for ratings of baseline voice quality fell in the moderate agreement range ($\kappa = 0.4$ to 0.6; Capozzoli, McSweeney & Sinha, 1999).

Table 3

The frequencies of unanimous and partial-agreement in swallow screening ratings by speech-language pathologist and nurse judges during the rating of abnormal clinical signs from movies of patients performing a brief swallow screening protocol.

Screening Question	S-LPs			RNs		Pooled across all 7 raters			
	Split decision (50% agreement)	*75% agreement*	***100% agreement***	*67% agreement*	***100% agreement***	57% agreement	*71% agreement*	*86% agreement*	***100% agreement***
Tongue lateralization	6%	22%	**72%**	33%	**67%**	13%	10%	30%	**48%**
Baseline voluntary cough	15%	30%	**55%**	48%	**53%**	25%	13%	33%	**30%**
Baseline voice quality	20%	29%	**51%**	58%	**42%**	25%	20%	30%	**25%**
Post swallow cough, throat clear or voice change	3%	22%	**75%**	25%	**75%**	10%	13%	33%	**45%**
Overall pass-fail result	13%	20%	**67%**	45%	**55%**	20%	25%	25%	**30%**

Table 4

Intra-class correlation and Cohen's Kappa statistics for inter-rater agreement (within, across and between professional groups) during the rating of abnormal clinical signs from movies of patients performing a brief swallow screening protocol.

Screening Question	Intra-class Correlation (S-LPs)			Intra-class Correlation (RNs)			Intra-class Correlation (Combined)			Cohen's Kappa for Inter-profession Agreement on Binary Dispositions
	-	95% Confidence Interval		-	95% Confidence Interval		-	95% Confidence Interval		
	Mean	*Lower Boundary*	*Upper Boundary*	*Mean*	*Lower Boundary*	*Upper Boundary*	*Mean*	*Lower Boundary*	*Upper Boundary*	
Tongue lateralization	0.84	0.73	0.91	0.53	0.21	0.74	0.84	0.75	0.91	0.66
Baseline voluntary cough	0.79	0.66	0.89	0.74	0.57	0.86	0.88	0.81	0.93	0.79
Baseline voice quality	0.78	0.64	0.88	0.45	0.07	0.69	0.81	0.70	0.89	0.43
Post swallow cough, throat clear or voice change	0.90	0.83	0.94	0.86	0.76	0.92	0.91	0.86	0.95	0.65
Overall pass-fail result	0.73	0.55	0.85	0.69	0.48	0.83	0.84	0.74	0.91	0.64

In total, 14 participants were determined to show penetration-aspiration scores of 3 or higher on at least one of the swallowing tasks in the screening protocol. Dysphagia was judged to be present in 24 participants. The results of the cross-tabulation of the screening movie and VF disposition scores are shown by profession in Table 5. None of the test results in this study met the a *priori* criterion that we had specified for adequate predictive power (the combination of sensitivity, specificity and negative predictive value scores > 0.6, a false negative rate < 0.2 and a positive likelihood ratio > 1.0).

The diagnostic utility of the "other signs of difficulty" question was not specifically tested in a cross-tabulation, due to the fact that the question was open ended. However, the items reported under this parameter were interesting: 13 patients were described to exhibit delayed swallow or prolonged transit times; 14 were described to need multiple swallows per bolus; 7 were described to exhibit bolus control problems with associated anterior spill or drooling. On a percentage basis, there were no obvious differences in the frequencies of specific types of comments (delay; multiple swallows, bolus control) between the S-LP and RN raters. Other comments, describing 1-3 patients per condition, included observations regarding breathing, behaviors (grimacing, impulsivity, obvious effort to swallow) and visible facial paresis. Figure 1 shows one such example, in which facial asymmetry was noticeable in the screening movie.

Table 5

The predictive utility of abnormal clinical signs for detecting aspiration and dysphagia, when observed by speech-language pathologists and nurses for patients performing a brief swallow screening protocol, and validated against concurrent videofluoroscopy.

Validation	Rater Group	Measure	Sensitivity	Specificity	Negative Predictive Value	Positive Predictive Value	False Positive Rate	False Negative Rate	Positive Likelihood Ratio
Penetration-Aspiration on VFSS	S-LP	Abnormal Baseline Tongue Lateralization	14%	72%	60%	22%	78%	40%	0.51
		Abnormal Baseline Voluntary Cough	36%	60%	63%	33%	67%	38%	0.89
		Abnormal Baseline Voice Quality	57%	56%	70%	42%	58%	30%	1.30
		Post Swallow Cough, Throat Clear or Voice Change	21%	52%	54%	20%	80%	46%	0.45
		Overall Screen Result of Fail	64%	16%	44%	30%	70%	56%	0.77
	RN	Abnormal Baseline Tongue Lateralization	7%	80%	61%	17%	83%	39%	0.36
		Abnormal Baseline Voluntary Cough	36%	80%	69%	50%	50%	31%	1.79
		Abnormal Baseline Voice Quality	50%	64%	70%	44%	56%	30%	1.39
		Post Swallow Cough, Throat Clear or Voice Change	50%	52%	65%	37%	63%	35%	1.04
		Overall Screen Result of Fail	57%	44%	65%	36%	64%	35%	1.02
Dysphagia on VFSS	S-LP	Abnormal Baseline Tongue Lateralization	21%	73%	37%	56%	44%	63%	0.78
		Abnormal Baseline Voluntary Cough	42%	67%	42%	67%	33%	58%	1.25
		Abnormal Baseline Voice Quality	58%	67%	50%	74%	26%	50%	1.75
		Post Swallow Cough, Throat Clear or Voice Change	38%	60%	38%	60%	40%	63%	0.94
		Overall Screen Result of Fail	71%	13%	22%	57%	43%	78%	0.82
	RN	Abnormal Baseline Tongue Lateralization	13%	80%	36%	50%	50%	64%	0.63
		Abnormal Baseline Voluntary Cough	25%	73%	38%	60%	40%	62%	0.94
		Abnormal Baseline Voice Quality	54%	80%	52%	81%	19%	48%	2.71
		Post Swallow Cough, Throat Clear or Voice Change	46%	47%	35%	58%	42%	65%	0.86
		Overall Screen Result of Fail	58%	47%	41%	64%	36%	59%	1.09

DISCUSSION

This study provides greater detail regarding the predictive power of individual clinical signs for detecting aspiration and dysphagia using a brief swallow screening protocol. Our results identified differences in the power of particular indicators, when judged by RNs compared to S-LPs. Overall, predictive power measures were poor compared to prior studies of longer screening protocols in which a separate (indirect) instrumental gold standard test has been used for validation (see Table 2). These results were not what we had expected and prompted us to consider a variety of explanations. The most obvious explanation is the fact that our study involved a direct comparison of clinical judgments and blinded VF ratings for exactly the same swallows. This is, to our knowledge, the first study to report such a direct comparison. In this section, we review the details of our study, its limitations, and several other possible explanations for differences between our study results and those of previous studies in the swallowing literature (see Table 2).

ITEM ANALYSIS

Of the five questions that were asked during the swallow screening movie rating, the observation of abnormal tongue lateralization had the lowest sensitivity, both for aspiration and for dysphagia detection. This clinical sign was not one that judges had difficulty agreeing on, either within or across professions, although the intra-class correlation amongst RN judges was only 0.53. When considered in isolation, the identification of abnormal tongue lateralization led to excessively high false positive rates for the identification of aspiration and chance-performance for identifying dysphagia. On this basis, we would argue that difficulty in lateralizing the tongue has questionable validity as a clinical sign of dysphagia and should not be interpreted as an indication of aspiration.

Abnormalities in the ability to produce a voluntary cough at baseline showed poor sensitivity for detecting aspiration or dysphagia according to the ratings of both professional groups, although intra-class correlations for inter-rater agreement were fairly strong. Nurse judgments of this sign showed good specificity (80%) and low false-positives (31%), leading to a positive likelihood ratio of 1.79 for aspiration. The speech-language pathology ratings of this sign showed modest specificity (67%), a low false-positive rate (33%) and a positive likelihood ratio of 1.25 for dysphagia. These findings suggest that this parameter may provide useful swallow screening data when combined with other indicators.

The validity of abnormal voice quality, either at baseline or post-swallow, has received extensive discussion elsewhere in the literature (Groves-Wright, Boyce & Kelchner, 2010; Warms & Richards, 2000; Waito et al., 2010). In this study, S-LP judges showed better agreement (ICCs = 0.78) than RNs (0.45) for perceptual judgments of baseline voice quality. This is probably not surprising, given the focus on perceptual voice assessment in speech-language pathology training programs. When all raters were pooled as a single rating group, consensus on this indicator was poor. Nonetheless, specificities for aspiration were modest for both rater groups with positive likelihood ratios of 1.3-1.39. Interestingly, this clinical sign had the strongest performance, in terms of likelihood ratios, for detecting dysphagia. Specificities for this sign were reasonably strong, consistent with previous findings in the literature (Waito et al., 2010). When RNs determined that baseline voice quality was abnormal, the patient was 2.71 times more likely to have dysphagia, as per the blinded videofluoroscopy ratings. These results support the continued consideration of baseline voice quality in swallow screenings, noting, however, that perceptual judgments of this feature may suffer from high inter-rater variability.

The observation of a post-swallow cough, throat clear, or change in voice quality was agreed upon most easily by judges. However, sensitivity and specificity of this measure for aspiration were poor, and false positives were high, particularly in the S-LP ratings. This sign yielded essentially equivocal findings regarding the possibility of underlying dysphagia. These results suggest that observations of coughing and voice quality post-swallow do not lead to accurate impressions regarding the presence of aspiration or dysphagia.

Finally, we asked judges to give an overall impression rating regarding the pass-fail status of a patient on the entire swallow screening protocol. This was not a score that was derived based on the results of the other screening questions, but an overall rating indicating the judge's opinion about whether the patient would require a referral for more detailed swallowing assessment, if they were seen in a typical emergency-room screening context. Agreement on this item was reasonably good, although RNs displayed greater variation than S-LPs. This parameter showed poor specificity for aspiration and dysphagia based on the S-LP ratings, and equivocal results for the RNs. The false negative rates on this question are somewhat alarming, particularly with respect to identifying dysphagia. Interestingly, according to this overall impression question, patients were more likely to be incorrectly classified as not having dysphagia by a speech-language pathologist than by a nurse!

DIFFERENCES IN SCREENING PROTOCOLS

As mentioned previously, the specific tasks in our brief swallow screening protocol were chosen based on a review of the swallow screening literature, and most closely resemble part 2 of the protocol used at the Massachusetts General Hospital (MGH-SST; Cohen, 2008, 2009). Our study did not consider the relationship between the thin liquid swallows examined in the swallow screening sequence and swallows of both liquid and other stimuli that may have occurred over the course of the entire videofluoroscopy protocol. We acknowledge that the choice to perform our screening protocol in a time-linked manner in the videofluoroscopy suite imposed a constraint on the number of swallows elicited in the screening test. From a mathematical perspective, the reduced number of swallows used in our study to determine true aspiration and dysphagia status would be most likely to lead to lower sensitivity statistics, in comparison to studies in which all swallows from a subsequent instrumental exam have been used. The observed result seems to confirm this expectation.

SAMPLE CONSIDERATIONS

Previous studies in which blinded validation of swallow screening protocols has been reported have included validation samples ranging in size from 50 to 100 participants (Clave et al., 2008; Cohen, 2008, 2009; Daniels et al., 1998; Martino et al., 2009a; Trapl et al., 2007). Our sample size, with 40 participants, was slightly smaller, but comparable to the validation sample sizes of 50-70 in the studies by Martino et al. (2009a), Daniels et al. (1998), and Trapl et al. (2007), with the important qualification that we studied a heterogeneous sample rather than the stroke population evaluated in all three of these prior studies.

With the exception of likelihood ratio measures, it is probably not appropriate to compare predictive power statistics for a protocol tested in a specific group to one tested in a heterogeneous sample because the prevalence of the underlying disorder in the sampled population may differ (Schoenfeld, 2009). Although swallow screening protocols have been most rigorously tested in etiologically specific groups (most commonly stroke) in the literature (e.g. Daniels et al., 2008; Martino et al., 2009a; Trapl, et al., 2007), it is fairly common practice to use swallow screening protocols in broader patient populations (e.g., Suiter and Leder, 2008). It was in this general context that we wanted to explore our questions regarding swallow screening utility. In retrospect, it is unfortunate that we did not collect additional diagnostic and etiological information regarding our participants, which might have permitted a subgroup analysis of those with stroke, brain injury or neurologic disease compared to those with unknown or non-neurogenic etiologies. However, to do such a subgroup analysis with adequate statistical power would have required a much larger sample size. Heterogeneity in the etiology of our participants may have led to increased sensitivities and lower specificities than studies of more etiologically-specific samples. As a rule, sensitivity results in our study were considerably lower than those reported in studies limited to the stroke population (Martino et al., 2009a; Perry, 2001; Massey & Jedlicka, 2002; Cichero, Heaton & Bassett, 2009; Daniels, et al., 2008; Trapl, et al., 2007). On the other hand, our sample comprised individuals referred for the investigation of suspected swallowing disorders. This meets one of the key criteria outlined by Sackett and Haynes for validity of a test, namely "independent, blind comparison of test results with a reference standard among a consecutive series of patients suspected (but not known) to have the target disorder" (Sackett & Haynes, 2002). To use a referred sample like this in a test validation study risks inflating the estimate of the true occurrence of the target underlying disorder, leading to reduced sensitivity and heightened specificity compared to studies of all-comers. Both of the prior studies in which swallow screening has been evaluated in heterogeneous populations appear to have been done in referred samples (Clave et al., 2008; Suiter & Leder, 2008). The sensitivities (and sample sizes) in our study were lower than reported in these studies, but specificities were comparable.

DATA COLLECTION PROCEDURES

Although our method of collecting time-linked data permitted a direct comparison across screening and videofluoroscopic data, the use of movies rather than live observation may well have altered the information that was considered by raters for the screening procedure. Screening procedures are supposed to be simple and transparent. In order to minimize the possibility of variable screening decisions across clinicians, such decisions should be made based on clinical signs that are directly evaluated and explicitly considered, rather than other contextual information which may be apparent. These requirements lay behind our method of posing specific and direct questions during the movie rating portion of our study. Nonetheless, qualitative review of the comments provided by raters, particularly regarding other signs of concern noted during the review suggests that both S-LP and nursing raters were attuned to contextual information beyond that which was queried in the screening protocol questions. This finding suggests that the predictive power reported in previous studies where the decisions have not been derived through direct, itemized questions may involve an unknown contribution of context.

RATING PROCEDURES AND DEFINITIONS

In reviewing the previous literature on screening, we found it quite difficult to ascertain the operational definitions used to define the target abnormalities of interest. Definitions for penetration-aspiration and dysphagia are lacking in many previous studies, as are the methodological details necessary to support replication. It is common to rate a wide number of physiological parameters on videofluoroscopy, ranging from lip and soft-palate closure to aspiration (Martin-Harris et al., 2008). In this study, we decided that the target underlying disorder of dysphagia would be defined as an abnormality resulting in a functional consequence (i.e., penetration-aspiration or pharyngeal residues). Our methods of classifying true disposition from the videofluoroscopy were intentionally set with low thresholds for failure (any single occurrence of a penetration-aspiration score ≥ 3 and/ or any single occurrence of a cumulative residue score ≥ 2). It is possible that our definition for dysphagia was more focused and less inclusive than that in other studies, where the criterion for a disposition of dysphagia has been reported as "any abnormality on videofluoroscopy" (Daniels et al., 1998; Martino et al., 2009). If so, one would expect higher sensitivities and lower specificities to be found. On the other hand, our threshold for classifying the screening results as abnormal was intentionally set to err on the side of identifying a problem in the event that at least 50% of the raters within a professional group queried the presence of an abnormality. This low-threshold definition also runs the risk of inflated sensitivity, reduced specificity and increasing false-negatives in validation against a gold-standard method like videofluoroscopy. We found both lower sensitivities and lower specificities with higher false negatives compared to previous studies where the classification threshold definitions are not transparently reported. This suggests the possibility that *both* our clinical screening and our VF classification thresholds may have been set too broadly, and that narrower definitions might need to be established.

POSSIBLE LIMITATIONS OF DIRECT COMPARISON

As noted in our introduction, prior studies of swallow screening performance have compared the single pass-fail result from a screen to the subsequent occurrence of problems, anywhere during the course of an instrumental examination. In our study, we took the approach of limiting our focus only to the limited number of thin liquid swallows in the screening protocol. It is possible that this direct focus led us to miss the power of the screening protocol to predict problems on a more comprehensive instrumental examination. In such a case, the result would have been increased sensitivity and reduced specificity. The fact that we saw reductions in both of these metrics could be interpreted as an argument against this possibility. It is our opinion that swallowing performance on a thin liquid screening challenge is not logically likely to predict true performance on different tasks, and that to draw such connections is to overstate the power of the original screening test. It should be remembered that the purpose of swallow screening tests is to obtain an accurate initial impression of swallowing performance, which should determine the appropriateness of sending a patient for further assessment. Water or thin liquids are arguably the most likely stimuli to elicit aspiration, and are therefore a good choice for items that are most likely to identify patients who require additional assessment. The emphasis on developing screening protocols that are also good predictors of dysphagia is interesting to consider in this respect. Our experience with this study has caused us to reflect that the five questions included in our study were thematically more oriented to detecting possible aspiration than to other aspects of dysphagia like residue. Indeed, we did not ask patients whether they had a sensation of residue, nor did we ask judges to comment specifically on any observations that might suggest residue (like multiple swallows per bolus). Based on these reflections, we would argue that it is theoretically improbable that the observation of tongue mobility, cough, and voice quality before and after swallows of water would be good predictors of post-swallow residue, either with water or with other stimuli.

CONCLUSIONS

It is difficult to compare results across different screening validation studies in the literature because methods of determining the disposition of dysphagia or aspiration have not been clearly described, sample inclusion criteria differ, and blinding has not always been used. In this study, we performed a direct (blinded) comparison of screening judgments to concurrent VF ratings, removing concerns that the screening and validation tests might capture variable physiological performance in a patient due to circumstantial differences and the passage of time. We had expected to see reasonably good correspondence between the observation of abnormal clinical signs and the true occurrence of aspiration and dysphagia. Our findings, therefore, came as a surprise and have caused us to reflect critically on the goals of swallow screening as a process. Based on our study, we have to conclude that swallow screening decisions based on a series of 3-4 thin liquid swallows do not have good clinical utility for detecting penetration-aspiration or dysphagia with thin liquid stimuli. In

particular, judgments regarding tongue lateralization and post-swallow cough or voice quality changes were not found to have predictive value.

Our results suggest that swallow screening results are imperfect, and suffer both from over-identification and under-identification of the underlying target disorders. In particular, the screening protocol tested in this study yielded a high number of false negative decisions, suggesting that patients with underlying dysphagia or penetration-aspiration would not have been detected by clinicians applying the screening. This finding is of concern, given the objective of swallow screening programs to accurately identify patients at risk for dysphagia and its consequences. Of the five focused questions that were asked during the rating of the screening movies, only the observations of abnormal baseline voice quality and voluntary cough showed acceptable predictive performance for identifying aspiration and dysphagia.

The findings are challenging to consider in terms of clinical practice recommendations. We suggest that clinicians need to recognize that swallow screenings are a very preliminary first step test towards identifying swallowing problems, and that they need to be treated as such. In this respect, we come back to the original expectations regarding screening put forward by the World Health Organization: "Screening tests sort out apparently well persons who probably have a disease from those who probably do not. A screening test is not intended to be diagnostic" (*CCI Conference on preventive aspects of chronic disease,* 1951). However, we believe that the current results also suggest that swallow screening tests may need to involve a greater number of swallows, or be repeated on more than one occasion over the first few days of a patient's admission to hospital if the goal is not to miss patients who may have dysphagia. Furthermore, our results point to the importance of following up with clinical and instrumental swallowing assessments in a timely manner in patients who are identified through swallow screening protocols. This will provide greater detail to inform patient management and should help to limit the over-zealous use of diet restrictions in patients whose swallow screening results are false positives. Certainly, our results reinforce the importance of recognizing that swallow screenings do not provide sufficient information for patient management and cannot be used to replace more detailed swallowing assessment in patients at risk for dysphagia.

REFERENCES

ASSiST: Acute Swallow Screen in Stroke and TIA. (2009). Retrieved June 24, 2010 from http://www.ircst.health.nsw.gov.au/__data/assets/pdf_file/0006/95685/Can_i_ASSIST_you_with_that.pdf

Bours, G. J., Speyer, R., Lemmens, J., Limburg, M. & de Wit, R. (2009). Bedside screening tests vs. videofluoroscopy or fibreoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review. *Journal of Advanced Nursing, 65*(3), 477-493.

Bravata, D. M., Daggett, V. S., Woodward-Hagg, H., Damush, T., Plue, L., Russell, S. et al., (2009). Comparison of two approaches to screen for dysphagia among acute ischemic stroke patients: Nursing admission screening tool versus National Institutes of Health Stroke Scale. *Journal of Rehabilitation Research and Development, 46*(9), 1127-1134.

Canadian Stroke Network (2005). SCORE (*Stroke Canada Optimization of Rehabilitation through Evidence*). *Post-Stroke Evidence-Based Recommendations: Screening for risk of pressure ulcers, falls, dysphagia, cognitive disorders and depression*. Retrieved June 24, 2010 from http://www.trilliumhealthcentre.org/west_GTA_stroke_network/professional/SCORE.pdf

Capozzoli, M., McSweeney, L., and Sinha, D. (1999). Beyond Kappa: A review of interrater agreement measures. *The Canadian Journal of Statistics, 27*(1), 3 – 23.

CCI Conference on preventive aspects of chronic disease. (1951). Retrieved June 24, 2010. from *http://whqlibdoc.who.int/php/WHO_PHP_34.pdf*

Cichero, J. A., Heaton, S. & Bassett, L. (2009). Triaging dysphagia: nurse screening for dysphagia in an acute hospital. *Journal of Clinical Nursing, 18*(11), 1649-1659.

Clave, P., Arreola, V., Romea, M., Medina, L., Palomera, E. & Serra-Prat, M. (2008). Accuracy of the volume-viscosity swallow test for clinical screening of oropharyngeal dysphagia and aspiration. *Clinical Nutrition, 27*(6), 806-815.

Cohen, A. K. (2008). MGH- Swallow Screening Tool (MGH-SST): Validation and implementation in acute neuro patients. Retrieved June 24, 2010, from http://www.docstoc.com/docs/21656737/MGH--Swallow-Screening-Tool-(MGH-SST)-Validation-and

Cohen, A. K. (2009). Creating a swallow screening program at Mass General Hospital: A model for development and implementation. *Perspectives on Swallowing and Swallowing Disorders, 18*(4), 123-128.

Daniels, S.K., Brailey, K., Priestly, D.H., Herrington, L.R., Weisberg, L.A. & Foundas, A.L. (1998). Aspiration in patients with acute stroke. *Arch Phys. Med Rehabil, 79*(1), 14-19.

Eisenhuber, E., Schima, W., Stadler, A., Schober, E., Schibany, N., Denk, D. M. et al. (2005). Evaluation of swallowing disorders with videofluoroscopy in Austria: a survey. *European Journal of Radiology, 53*(1), 120-124.

Grey-Bruce Health Network Dysphagia Screening Tool (2008). Retrieved June 24, 2010, from http://www.gbhn.ca/ebc/documents/DYSPHAGIASCREENINGTOOLJuly2008.pdf

Groves-Wright , K. J., Boyce, S. & Kelchner, L. (2010). Perception of wet vocal quality in indentifying penetration/aspiration during swallowing. *Journal of Speech, Language and Hearing Research, 53*(3), 620-632.

Iredell Dysphagia Screen (2006). Retrieved June 24, 2010, from http://www.encsn.org/Hospitals%20Plans%20of%20Care%20Templates/Iredell.dysphagia.screen.09.pdf

Joint Commission on the Accreditation of Healthcare Organizations (2004). *Stroke Performance Measurement Implementation Guide.* Retrieved June 24, 2010 from http://www.jointcommission.org/CertificationPrograms/PrimaryStrokeCenters/implementation_guide.htm

Leder, S. B., Suiter, D. M. & Green, B. G. (2010). Silent Aspiration Risk is Volume-dependent. *Dysphagia, online first.* doi: 10.1007/s00455-010-9312-2

Logemann, J. A. (1993). *Manual for the videofluoroscopic study of swallowing.* 2nd edition. Austin, TX: Pro-Ed.

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

Lothian Dysphagia Screening Test (2005). Retrieved June 24, 2010, from http://www.lothianstrokemcn.scot.nhs.uk/publications/documents/dysphagiascreeningtestv4a_270705.pdf

Martin-Harris, B., Brodsky, M. B., Michel, Y., Castell, D. O., Schleicher, M., Sandidge, J. et al. (2008). MBS measurement tool for swallow impairment--MBSImp: establishing a standard. *Dysphagia, 23*(4), 392-405.

Martino, R., Pron, G. & Diamant, N.E. (2000). Screening for oropharyngeal dysphagia in stroke: Insufficient evidence for guidelines. *Dysphagia 15*(1), 19-30.

Martino, R., Silver, F., Teasell, R., Bayley, M., Nicholson, G., Streiner, D. L. et al. (2009a). The Toronto Bedside Swallowing Screening Test (TOR-BSST): development and validation of a dysphagia screening tool for patients with stroke. *Stroke, 40*(2), 555-561.

Martino, R., Streiner, D. L., Maki. E., Diamant, N. (2009b). A sensitivity

analysis to determine whether ten teaspoons of water are really necessary. *Dysphagia, 24*(4),473.

Massey, R. & Jedlicka, D.(2002). The Massey bedside swallowing screening. *Journal of Neuroscience Nursing, 34*(5), 252-260.

McCullough, G. H., Rosenbek, J. C., Wertz, R. T., McCoy, S., Mann, G. & McCullough, K. (2005). Utility of clinical swallowing examination measures for detecting aspiration post-stroke. *Journal of Speech-Language Hearing Research, 48*(6), 1280-1293.

Oklahoma Dysphagia Screening Tool (n.d.). Retrieved June 24, 2010, from http://www.integris-health.com/NR/rdonlyres/22E4D1AA-E0D6-4F78-9AC5-C49ED00993A4/0/Dysphagia_Screening.pdf

Perry, L. (2001). Screening swallowing function of patients with acute stroke. Part II: Detailed evaluation of the tool used by nurses. *Journal of Clinical Nursing, 10*, 474-481.

Perry, L. & Love, C. P. (2001). Screening for dysphagia and aspiration in acute stroke: a systematic review. *Dysphagia 16*(1), 7-18.

Rosenbek, J.C., Robbins, J.A., Roecker, E.B., Coyle, J.L. & Wood, J.L. (1996). A penetration-aspiration scale. *Dysphagia 11*(2), 93-98.

Sackett, D.L. & Haynes, R.B. (2002). The architecture of diagnostic research. *British Medical Journal, 324*(7336), 539-541.

Sackett, D.L., Straus, S.E., Richardson, W.S., Rosenberg, W. & Haynes, R.B. (2000). *Evidence-Based Medicine: How to practice and teach EBM*. Toronto, ON: Churchill Livingstone

Scherokman, B. (1997). Selecting and interpreting diagnostic tests. *The Permanente Journal, 1*(2), 1-7.

Schoenfeld, P. (2009). Is this diagnostic test useful? *Am J Gastroenterol, 104*(9), 2139-2142.

Scottish Intercollegiate Guidelines Network (2004). *Management of patients with stroke: Identification and management of dysphagia.* Retrieved June 24, 2010, from http://www.sign.ac.uk/pdf/sign78.pdf

St. George Dysphagia Screening Tool (n.d.). Retrieved June 24, 2010, from http://www.stgeorgedgp.asn.au/division-programs/uploadedFiles/1236656727857-1959.pdf

Suiter, D. M. & Leder, S. B. (2008). Clinical utility of the 3-ounce water swallow test. *Dysphagia, 23*(3), 244-250.

Trapl, M., Enderle, P., Nowotny, M., Teuschl, Y., Matz, K., Dachenhausen, A. et al. (2007). Dysphagia bedside screening for acute-stroke patients: the Gugging Swallowing Screen. *Stroke, 38*(11), 2948-2952.

Waito, A., Bailey, G. L., Molfenter, S. M., Zoratto, D. C. & Steele, C. M. (2010). Voice-quality Abnormalities as a Sign of Dysphagia: Validation against Acoustic and Videofluoroscopic Data. *Dysphagia, online first.* doi: 10.1007/s00455-010-9282-4

Warms, T. & Richards, J. (2000). "Wet Voice" as a predictor of penetration and aspiration in oropharyngeal dysphagia. *Dysphagia, 15*(2), 84-88.

Zoratto, D. C., Chau, T. & Steele, C. M. (2010). Hyolaryngeal excursion as the physiological source of swallowing accelerometry signals. *Physiological Measures, 31*(6), 843-855.

ACKNOWLEDGMENTS

This study was funded by a grant to the first author from the Ontario Centres of Excellence, Ontario Ministry of Research and Training, with industry support from Panacis Medical Inc. The first author also acknowledges support from a Canadian Institutes of Health Research New Investigator Award (Grant #69521). The authors acknowledge the support of Toronto Rehabilitation Institute which receives funding under the Provincial Rehabilitation Research Program from the Ministry of Health and Long-term Care in Ontario. The views expressed are those of the authors and do not reflect those of the ministry.

The authors would like to acknowledge the assistance of Janice Bennett, Nicole Buesselberg, Joon Lee, Mitsuko Takeuchi, Ka Lun Tam and Annemarie Weeda with data collection and processing.

AUTHOR'S NOTE

Correspondence should be sent to Catriona M. Steele, Ph.D., Senior Scientist, Toronto Rehabilitation Institute, 550 University Avenue, 12th floor (research) Toronto, ON., M5G 2A2, Canada. E-mail: steele.catriona@torontorehab.on.ca

Received date : August 25, 2010
Accepted date : February 14, 2011