JAN JOURNAL OF ADVANCED NURSING

REVIEW PAPER

Bedside screening tests vs. videofluoroscopy or fibreoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review

Gerrie J.J.W. Bours, Renée Speyer, Jessie Lemmens, Martien Limburg & Rianne de Wit

Accepted for publication 31 October 2008

Correspondence to G.J.J.W. Bours:
e-mail: g.bours@zw.unimaas.nl

Gerrie J.J.W. Bours PhD RN
Assistant Professor
Department of Health Care Studies,
Maastricht University, School for Public Health and Primary Care,
and Researcher,
University Zuyd, Expertise Centre on Autonomy and Participation,
The Netherlands

Renée Speyer PhD SLT
Senior Researcher
Department of O.R.L. and Head & Neck Surgery, University Hospital Maastricht,
The Netherlands

Jessie Lemmens MSc SLT
Researcher
Department of Health Care Studies,
Maastricht University, School for Public Health and Primary Care,
and Researcher,
University Zuyd, Expertise Centre on Autonomy and Participation,
The Netherlands

Martien Limburg PhD MD
Professor of Neurology
Department of Neurology, University Hospital Maastricht,
The Netherlands

continued on page 478

BOURS G.J.J.W., 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
doi: 10.1111/j.1365-2648.2008.04915.x

Abstract

Title. Bedside screening tests vs. videofluoroscopy or fibreoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review.

Aim. This paper is a report of a systematic review conducted to determine the effectiveness and feasibility of bedside screening methods for detecting dysphagia in patients with neurological disorders.

Background. Dyspaghia affects 22–65% of patients with neurological conditions. Although there is a large variety of bedside tests to detect dysphagia, it is unknown which have the best psychometric properties and are feasible for nurses to use.

Data sources and review methods. An electronic database search was carried out using Medline (PubMed), Embase, CINAHL, and PsychLit, including all hits up to July 2008. The search terms were dysphagia, sensitivity, specificity, diagnosis, and screening. The methodological quality of included studies was assessed.

Results. Thirty-five out of 407 studies were included in the review. Eleven studies with sufficient methodological quality revealed that trial swallow tests using water had sensitivities between 27% and 85% and specificities between 63% and 88%. Trial swallow tests with different viscosities led to sensitivities ranging from 41% to 100% and specificities of 57% to 82%. Combining water tests with oxygen desaturation led to sensitivities between 73% and 98% and specificities between 63% and 76%. Single clinical features, such as abnormal gag, generally had low sensitivity and specificity.

Conclusion. A water test combined with pulse oximetry using coughing, choking and voice alteration as endpoints is currently the best method to screen patients with neurological disorders for dysphagia. Further research is needed to establish the most effective standardized administration procedure for such a water test, and to assess the value of pulse oximetry, in addition to a trial swallow to detect silent aspiration.

Keywords: dysphagia, fibreoptic endoscopic, neurological disorders, screening tests, swallowing, systematic review, videofluoroscopy

Rianne de Wit PhD
Professor of Nursing Science
Department of Health Care Studies, Section of Nursing Science, Maastricht University, and University Hospital Maastricht, School for Public Health and Primary Care, The Netherlands

Introduction

Dysphagia is a very common feature of neurological disorders. It has been reported to affect 22–65% of patients with acute stroke (Daniels *et al.* 1998, Smithard *et al.* 1998) 36% of symptomatic patients with Parkinson's disease (Mari *et al.* 1997), and more than 30% of those with multiple sclerosis (Prosiegel *et al.* 2004). Aspiration is one of the most critical signs of oropharyngeal dysphagia, and may lead to chest infection, malnutrition, prolonged hospital stay and mortality (Smithard *et al.* 1996). Approximately one-third of patients with dyspaghia develop pneumonia requiring treatment (Agency for Health Care Policy and Research 1999, Mann *et al.* 1999, Smith Hammond & Goldstein 2006). To ensure safe, high quality care, it is desirable that the identification and management of dysphagia is incorporated in an appropriate risk management policy as part of clinical governance arrangements.

Clinical screening is important to identify patients who aspirate from the overall population with oropharyngeal dysphagia and to initiate early referral for diagnosis and treatment to minimize health risks. Videofluoroscopic (VF) evaluation is often considered the gold standard for assessing dysphagia (Rao *et al.* 2003). This technique provides dynamic imaging of the swallowing function by visualizing the bolus during the process of swallowing. Adding a contrast material such as barium sulphate allows the bolus to be followed as it travels through the alimentary tract (Murray 1999a). However, the radiation exposure in VF makes frequent test repetitions inappropriate. Fibreoptic endoscopic evaluation of swallowing (FEES) is safe and well tolerated, and has been found to be just as valuable as VF in diagnosing dysphagia (Langmore *et al.* 1991, Rao *et al.* 2003). FEES examination requires the transnasal passage of a flexible laryngoscope into the hypopharynx, whereupon food and liquid can be presented and the swallowing activity can be videotaped (Murray 1999b). It does, however, require a skilled operator and specialized equipment.

The complexity of symptoms and lack of specificity of the underlying disease that causes dysphagia necessitate a multidisciplinary approach, in which patient management and therapeutic options are discussed. Within this multidisciplinary team, nurses have an important role in assessing dysphagia, observing symptoms and reactions, using methods to relieve it and evaluating the effects. In fact, impaired swallowing is a nursing diagnosis accepted by the North American Nursing Diagnosis Association (NANDA 2007). Therefore, a feasible method for screening is needed that can easily be used by nurses to decide whether a patient can be given anything by mouth and to minimize unnecessary restriction of oral intake.

Although there are currently many bedside tests, it is not clear which of these have the best psychometric and feasible properties and are easy to administer. The challenge is to construct a screening tool which can easily be taught, is quickly administered and non-invasive, causes no distress to patients and produces reliable results. Some reviews have been performed previously, but these were not systematic, disregarded the quality of the studies or were limited to patients with stroke (Ott & Pikna 1993, Lambert & Gisel 1996, Martino *et al.* 2000, Perry & Love 2001, Kalf 2002, Ramsey *et al.* 2003, Westergren 2006). Many neurological disorders, ranging from static conditions to progressive disorders, may affect the mechanism of swallowing, and screening of swallowing impairments in such conditions has been suggested to protect patients from adverse events. We were especially interested in screening measures applicable to this broad patient population, in order to develop a dysphagia screening protocol that could be implemented on all neurological wards. However, we excluded older adults with psychiatric conditions and patients with dysphagia caused by cancer. The former are difficult to instruct, and this may affect the results of screening or the application of the reference test. In the case of cancer of the head and neck region, dysphagia is often caused by an anatomical modification of the oral cavity and pharynx due to the cancer or its treatment. In this review, we were interested in bedside screening methods to detect dysphagia in patients with neurological disorders.

The review

Aims

The aim of the present systematic review was to determine the effectiveness of bedside methods to detect dysphagia in patients with neurological disorders. A secondary aim was to establish the practical feasibility of the screening methods. The following research question was formulated: which bedside screening method has the best psychometric and feasibility properties to detect dysphagia in patients with neurological disorders, compared with VF evaluation or FEES?

Design

We followed the steps suggested by the Cochrane Collaboration for reviewing the effects of intervention studies because, when we started this review, no guidelines were available from the Cochrane Collaboration for reviews on diagnostic test accuracy. Such guidelines were officially launched by the Cochrane collaboration in October, 2007, and a handbook for diagnostic test accuracy reviews is still under construction (Cochrane Collaboration 2008).

Search methods

We performed a computerized search using the databases Medline (Pubmed), Embase, CINAHL and PsychLit. All hits up to 26 June 2008 were used, and references in selected studies and citations of relevant reviews were checked for further references. The search terms used for each database and a flowchart of the abstracts identified is presented in Table 1.

Search outcome

The search strategy identified a total of 407 unique papers. Two reviewers (GB and RS) independently made an initial selection of the studies based on the abstracts. In doubtful cases, the entire paper was screened. Differences of opinion were resolved by discussion. The following inclusion and exclusion criteria were used.

The bedside screening test of interest had to be compared with a VF swallowing test or a FEES in a cross-sectional

Table 1 Search terms for the databases and flow chart of unique identified abstracts

Database	Search terms	Abstracts identified	Excluded
PubMed (medline and premedline)	a. Dysphagia ('Deglutition Disorders/diagnosis' [MeSH:NoExp] OR 'Deglutition Disorders/nursing' [MeSH:NoExp] OR 'Deglutition Disorders/prevention and control' [MeSH:NoExp])	225	
	b. Sensitiv*[Title/Abstract] OR sensitivity and specificity[MeSH Terms;exp] OR diagnos*[Title/Abstract] OR diagnosis[MeSH:noexp] OR diagnostic* [MeSH:noexp] OR diagnosis, differential (MeSH:noexp) OR diagnosis [Subheading:noexp] OR screening		
	c. #a AND #b		
Embase	a. Dysphagia /diagnoses in DEM, DER, DRM, DRR	22	
	b. Sensitivity and specificity/all subheadings in DEM, DER, DRM, DRR		
	c. #a and #b		
CINAHL	a. Swallowing – impairment – Saba – HHCC/all topical subheadings or impaired swallowing Nanda/all topical subheadings/ or deglutition disorder/all topical subheadings	36	
	b. Sensitivity and specificity/all topical subheadings or ROC curve/all topical subheadings		
	c. #a and #b		
PsychLit	a. Dysphagia in MJ, MN or pharyngeal disorders in MJ, MN or swallowing in MJ, MN	88	
	b. Psychometrics or sensitivity or specificity		
	c. #a and #b		
Total identified abstracts electronically		371	
Reference check		36	
Total identified abstracts		407	
No focus on screening			321
Inclusion criteria	1. Screening compared with videofluoroscopic or fibreoptic endoscopic evaluation of swallowing		
	2. Cross-sectional design or (randomized) clinical trial		
	3. Aspiration and/or penetration as endpoint		
	4. Predominant adult patients with neurological disorders		
	5. Non-invasive screening method		
	6. Written in English, German or Dutch		
Excluded papers			51
Included papers		35	

design or in a (randomized) clinical trial. The endpoint of the VF or FEES was defined as aspiration or aspiration and/or penetration. Aspiration is defined as the entry of material below the level of the true vocal folds, while penetration is defined as the entry of material in the laryngeal vestibule but remaining above the levels of the vocal folds (Murray 1999a,b). A study was also included if the screening method to be studied had not been applied at the bedside, but could be carried out by nurses at the bedside in the opinion of the two reviewers. The study population had to consist predominantly of adult patients with neurological disorders. If the screenings method was invasive, or only used to determine dysphagia in patients other than those with neurological disorders, such as patients with cancer or older patients with psychiatric conditions, the study was excluded. For practical reasons, only publications in English, German, or Dutch were included. The final outcome of all the searches consisted of 35 papers. Table 1 presents inclusion and exclusion details.

Quality appraisal

There is empirical evidence that diagnostic studies with methodological shortcomings overestimate the accuracy of a diagnostic test (Lijmer *et al.* 1999). Hence, study quality was assessed independently by two reviewers (GB and RS), using criteria adapted from a standard form for research on the accuracy of diagnostic tests developed by the Dutch Cochrane Centre (http://www.cochrane.nl). This form, which consists of nine criteria, requires the evaluator to allocate a plus to a criterion if the item has been addressed, a minus if the item has been violated, and a question mark if no information is available. A plus/minus was allocated to indicate that a criterion was partially satisfied. Criteria 1–6 refer to the validity of the studies, criterion 7 to their generalizability, and items 8 and 9 to their reliability. Table 2 presents an overview of these criteria.

Based on the criteria scored, we gave each study an overall quality rating in terms of validity, generalizability and reliability, using the qualifications 'sufficient', 'doubtful' and 'insufficient'. These qualifications were used as follows: a study was categorized as 'insufficient' when no data were available to calculate test characteristics (a minus on item 6). A study was categorized as 'sufficient' if no more than one item had been allocated a minus or question mark. If two items had been allocated either a minus or question mark, the study was considered to be 'doubtful', and if more than two items had been allocated either a minus or a question mark, the study was considered to have 'insufficient' validity, generalizability, and reliability. If an item received a plus/minus, it was given the benefit of doubt in the overall evaluation. Differences in quality ratings were resolved by discussion.

Results of the quality assessment are shown in Table 3. Eleven studies (31%) were assessed as having sufficient quality (Daniels *et al.* 1997, Mari *et al.* 1997, Smithard *et al.* 1998, Logemann *et al.* 1999, Smith *et al.* 2000, Lim *et al.* 2001, McCullough *et al.* 2001, Leder & Espinosa 2002, Mann 2002, Chong *et al.* 2003), 11 (31%) were of doubtful quality (Kidd *et al.* 1993, Garon *et al.* 1995, Collins & Bakheit 1997, Daniels *et al.* 1998, Addington *et al.* 1999, Koopman *et al.* 2004, Rosenbek *et al.* 2004, Shaw *et al.* 2004, Wu *et al.* 2004, Nishiwaki *et al.* 2005, Wang *et al.* 2005) and 13 studies (37%) were found to have insufficient quality (Linden & Siebens 1983, Horner & Massey 1988, Horner *et al.* 1988, 1993, Splaingard *et al.* 1988, DePippo *et al.* 1992, Zenner *et al.* 1995, Sherman *et al.* 1999, Warms & Richards 2000, Massey & Jedlicka 2002, Higo *et al.* 2003, Tohara *et al.* 2003, Lam *et al.* 2007). A description of the 11 studies with sufficient quality is presented in Table 4. The criterion that was satisfied by the smallest number of studies (25 of the 35 selected papers, and six of the 11 papers with good methodological quality) was that the index test had to

Table 2 Items for methodological assessment of the studies

Items	Description
1	Were the reference test and the index test interpreted independently (blind)?
2	Was the index test applied independent of relevant information on clinical data of the patient regarding the target condition?
3	Was the reference test applied to all patients who received the index test?
4	Was the period between the reference test and the index test short enough to be reasonably sure that the target condition did not change between the two tests? (within 24 hours in acute stroke, and within 7 days in other neurological diseases)
5	Was the selection of the study population valid?
6	Are data presented in enough detail to calculate appropriate test characteristics?
7	Was the study population appropriate to evaluate the proposed use of the index test?
8	Was the index test described in detail so it could be reproduced?
9	Were satisfactory definitions used for normal/abnormal reference test results and normal/abnormal index test results?

Table 3 Quality of the included studies

	Items									
	Validity						Generalizability	Reliability		
References	1	2	3	4	5	6	7	8	9	Conclusion
Chong *et al.* 2003	+	−	+	+	+	+	+	+	+	Sufficient
Daniels *et al.* 1997	+	+	+	−	+	+	+−	+	+	Sufficient
Leder and Espinosa 2002	+	−	+	+	+	+	+	+	+	Sufficient
Lim *et al.* 2001	+	+	+	+	+	+	+	+	+	Sufficient
Logemann *et al.* 1999	+	−	+	+	+	+	+−	+	+	Sufficient
Mann 2002	+	+	+	−	+	+	+	+	+	Sufficient
Mari *et al.* 1997	+	−	+	+	+	+	+	+	+	Sufficient
McCullough *et al.* 2001	+	+	+	+	+	+	+	+	+	Sufficient
Smith *et al.* 2000	+	+	+	+	+	+	+	−	+	Sufficient
Smithard *et al.* 1998	+	+	+	+	?	+	+	+	+	Sufficient
Trapl *et al.* 2007	+	−	+	+	+	+	+	+	+	Sufficient
Addington *et al.* 1999	+	+	+	?	−	+	+	+	+	Doubtful
Collins and Bakheit 1997	?	−	+	+	+	+	+	+	+	Doubtful
Daniels *et al.* 1998	+	?	+	?	+	+	+−	+	+	Doubtful
Garon *et al.* 1995	−	−	+	+	+	+	+−	+	+	Doubtful
Kidd *et al.* 1993	?	+	+	?	+	+	+	+	+	Doubtful
Koopman *et al.* 2004	+	−	+	+	+	+	+−	−	+	Doubtful
Nishiwaki *et al.* 2005	?	−	+	+	+	+	+	+	+	Doubtful
Rosenbek *et al.* 2004	+	?	+	+	?	+	+	+	+	Doubtful
Shaw *et al.* 2004	+	−	+	+	−	+	+−	+	+	Doubtful
Wang *et al.* 2005	+	−	+	+	?	+	+−	+	+	Doubtful
Wu *et al.* 2004	?	−	+	+	+	+	+−	+	+	Doubtful
DePippo *et al.* 1992	−	−	+	?	+	+	+	+	+	Insufficient
Higo *et al.* 2003	?	−	+	+	?	+	+−	+	+	Insufficient
Horner and Massey 1988	?	−	?	?	?	+	+	−	−	Insufficient
Horner *et al.* 1993	?	?	+	?	+	+	+	+	+	Insufficient
Horner *et al.* 1988	?	?	+	?	?	+	+	−	−	Insufficient
Lam *et al.* 2007	+	+	+	?	?	−	+	+	−	Insufficient
Linden & Siebens 1983	−	−	+	?	?	−	+	−	+	Insufficient
Massey and Jedlicka 2002	?	+	?	−	−	−	+	−	−	Insufficient
Sherman *et al.* 1999	+	−	+	+	−	−	+−	+	+	Insufficient
Splaingard *et al.* 1988	+	−	+	+	?	+	+−	−	−	Insufficient
Tohara *et al.* 2003	−	−	+	+	?	+	+	+	+	Insufficient
Warms and Richards 2000	+	−	+	+	+	−	+−	+	+	Insufficient
Zenner *et al.* 1995	−	−	+	−	?	+	+−	+	+	Insufficient

(1) Were the reference test and the index test interpreted independently (blind)?; (2) Was the index test applied independent of relevant information on clinical data of the patient regarding the target condition?; (3) Was the reference test applied to all patients who received the index test?; (4) Was the period between the reference test and the index test short enough to be reasonably sure that the target condition did not change between the two tests? (within 24 hours in acute stroke, and within 7 days in other neurological diseases); (5) Was the selection of the study population valid?; (6) Are data presented in enough detail to calculate appropriate test characteristics?; (7) Was the study population appropriate to evaluate the proposed use of the index test?; (8) Was the index test described in detail so it could be reproduced?; (9) Were satisfactory definitions used for normal/abnormal reference test results and normal/abnormal index test results?; conclusion, Are the results valid and applicable (sufficient: 8 or 9 items rated with a plus; doubtful: 7 items rated with a plus, and insufficient; 6 or less items rated with a plus or a minus on item 6)?; A plus/minus mark was rated as a plus. (+) The item has been addressed; (−) the item has been violated; (+−) the item has been partly addressed; (?) no information about the item was available in the paper.

be performed even though relevant information on clinical data were available (item 2), mostly because the patients had been referred for evaluation of swallowing difficulties. Fourteen studies were given a negative evaluation or failed to provide information about whether the reference test had been interpreted independently from the index (item 1), and 15 studies had a population selection method that was prone to bias (item 5). In addition, 13 papers failed to give the time interval between the index test and the reference test or did not mention it in an appropriate form (item 4).

Data abstraction

Original data from the 11 studies with sufficient quality were retrieved to calculate the prevalence, sensitivity, specificity, positive predictive value, negative predictive value, likelihood ratio of a positive test (LR+), and likelihood ratio of a negative test (LR−) (with 95% confidence intervals).

Statistical pooling proved impossible for various reasons, including the heterogeneity of the tests, differences in the way in which similar tests were implemented, or the use of different endpoints either in the reference test or index tests. The synthesis is therefore expressed by means of summary techniques, based only on the 11 studies that had sufficient methodological quality (Tables 5 and 6). We were hoping to find a bedside test with a high sensitivity (at least 70%) and at least a moderate specificity (at least 60%). The feasibility of a test was estimated in terms of the time required to apply it and its complexity (i.e. the need to use specialized equipment or a variety of materials).

Results

We first report on the characteristics of the included studies, then describe the diagnostic performance of the screening methods used in the 11 studies with sufficient quality, and finally report on the feasibility of these studies.

Characteristics of the included studies

Four of the 35 relevant studies used FEES as the reference test (Lim *et al.* 2001, Leder & Espinosa 2002, Chong *et al.* 2003, Trapl *et al.* 2007). All other studies used VF evaluation as a reference test and mentioned that the test material had been impregnated with barium. All studies used aspiration and/or penetration as the endpoint. A variety of bedside tests were used to determine aspiration, or aspiration and/or penetration. The majority were based on trial swallows, either using water in various aliquots or using liquids with a range of viscosities, thickened liquids, semi-solids and solids. Three papers described a bedside test which combined trial swallows and pulse oximetry (Smith *et al.* 2000, Lim *et al.* 2001, Chong *et al.* 2003). Four studies involved the use of pulse oximetry alone (Collins & Bakheit 1997, Sherman *et al.* 1999, Higo *et al.* 2003, Wang *et al.* 2005).

Three studies used a variety of clinical features to assess the risk of aspiration and/or penetration, such as abnormal gag, volitional cough, reduced laryngeal elevation, voice alteration, and the presence of dysphonia and dysarthria, using predefined cut-off points (Daniels *et al.* 1997, Logemann *et al.* 1999, McCullough *et al.* 2001). Other researchers examined the cough reflex elicited with acid dissolved in saline (Addington *et al.* 1999) or used cervical auscultation in addition to clinical examination (Zenner *et al.* 1995) and bronchial auscultation (Shaw *et al.* 2004). Some authors used medical history components to identify patients at risk of aspiration and/or penetration (McCullough *et al.* 2001, Rosenbek *et al.* 2004). Finally, four studies used a standardized form with a variety of clinical features, combined with a trial swallow test in which a specified cut-off point on the scoring list indicated patients at risk of aspiration and/or penetration (Logemann *et al.* 1999, Leder & Espinosa 2002, Mann 2002, Lam *et al.* 2007).

Diagnostic performance of the screening methods

The diagnostic performance of the 11 studies with sufficient methodological qualities is summarized in Tables 5 and 6, divided into seven categories.

Trial swallow using water

These tests used water, in various aliquots and administered in various ways (see also Table 4), as a trial swallow. Five studies used the water test to evaluate aspiration and/or penetration (Daniels *et al.* 1997, Mari *et al.* 1997, Smithard *et al.* 1998, Lim *et al.* 2001, Chong *et al.* 2003). Sensitivity ranged from 27% to 85%, while specificity ranged from 50% to 88%. Positive likelihood ratios ranged from 2·1 to 3·6, while negative likelihood ratios ranged from 0·2 to 0·8. Studies with more than one endpoint on the index test (coughing, choking and wet voice) as indicators of aspiration and/or penetration satisfied our predefined values of sensitivity and specificity, although the study by Smithard *et al.* (1998) was only the case when the risk of aspiration was assessed by a doctor. Lim *et al.* (Lim *et al.* 2001) and Chong *et al.* (2003) both gave patients 50 mL of water in 10 mL aliquots, while Smithard *et al.* (1998) started with a 5-mL spoonful of water and continued the test, if it was safe for the patient, by having them drink 60 mL of water in 2 minutes.

Trial swallow using different viscosities

Four studies used a range of liquids, semi-solids and solids to evaluate aspiration and/or penetration (Logemann *et al.* 1999, Smith *et al.* 2000, McCullough *et al.* 2001, Trapl *et al.* 2007). The aliquots and methods by which the trial swallow material was administered varied between the studies. Sensitivity of the trial swallow test ranged from 41% to 100% and specificity from 57% to 82%, depending on the indicators chosen as endpoints. Two studies (Logemann *et al.* 1999, McCullough *et al.* 2001) used single features to indicate the risk of aspiration, but satisfactory psychometric properties

Table 4 Characteristics of the included studies with sufficient methodological quality

References	Population	*n*	Reference test	Index test
Chong *et al.* 2003	Patients with recent stroke, or history of previous stroke who were referred to the Speech Therapy Department for the assessment of suspected dysphagia. Age ≥65 years	50	Type of test: FEES Swallow material: 3–5 spoonfuls with different food consistencies (honey-thickened, nectar-thickened, thin fluids and paste consistency which was colored with a blue food dye Endpoint: aspiration or penetration or if the speech therapist felt that it was unsafe to con tinue	Type of test: clinical aspiration test combined with pulse oximetry Swallow material: 50 mL water in 10 mL aliquots Endpoint: coughing or choking or changing voice quality on water test or a desaturation of ≥2% above baseline readings
Daniels *et al.* 1997	Male patients with new neurological deficit (stroke). Mean age 66 years (range 41–88)	59	Type of test: VF Swallow material: liquid barium from a cup or straw at volumes of 3, 5, 10, 20 mL, and ½ teaspoon barium paste, half of a cookie, and arbitrary amount of thinned liquid barium Endpoint: aspiration and/or penetration	Type of test: 1. Oropharyngeal examination: abnormal gag reflex (no or weakened response of velum or pharyngeal wall constriction unilaterally or bilaterally, on tactile stimulation of the poster ior pharyngeal wall); abnormal volitional cough (patient gives no response or a weak or vocalized response to cough on command); perceptual judgment of dysarthria (speech disorder resulting from disturbance in muscular control affecting the areas of respiration, articulation, phonation, resonance and/or prosody); perceptual judgment of dysphonia (disturbance in the parameters of vocal quality, pitch, or intensity), all rated binarily as present/absent 2. Clinical swallowing examination Swallow material: calibrated volumes of water starting with 5 mL liquid administered by cup or straw (preference patient) progressing to 10 and 20 mL and administered twice to a total of 70 mL Endpoint: cough after swallowing and/or voice change after swallow immediately or within 1 minute
Leder & Espinosa 2002	Stroke patients who were referred by neurology to SLP for swallowing evaluation. Mean age 70 years	49	Type of test: FEES Swallow material: 5 mL volume food boluses dyed with blue food coloring. First puree, fol lowed by liquid (milk) and then a solid cracker Endpoint: aspiration	Type of test: 1. Clinical testing, determining dysphonia, dysarthria, abnormal volitional cough, abnormal gag reflex. 2. Water test Swallow material: single water bolus with a straw Endpoint: cough after swallow and voice change after swallow (phonating ah). No aspiration was defined as ≤1 variable being identified on clinical testing and water test; aspiration was defined as ≥2 variables being identified

Table 4 (*Continued*)

References	Population	*n*	Reference test	Index test
Lim *et al.* (Lim *et al.* 2001)	Acute stroke patients admitted to a stroke unit. Mean age 67·5 years (SD 11·73)	50	Type of test: FEES Swallow material: three or more boluses of 5 mL of water thickened to a pudding consistency with thick and easy Endpoint: aspiration	Type of test: water test and pulse oximetry Swallow material: 50 mL water in 10 mL aliquots Endpoint: coughing or choking or a change in voice quality or a desaturation of ≥2%
Logemann *et al.* 1999	Patients referred for assessment of potential dysphagia. Main diagnoses was stroke, other diagnoses were treatment for head and neck cancer, spinal cord and other aetiologies. Mean age 65 years (14–97)	200	Type of test: VF Swallow material: modified barium swallow; no further information available Endpoint: aspiration, oral disorder, pharyngeal delay, pharyngeal disorder	Type of test: screening procedure consisted of 28 items divided into five categories, each including multiple variables: medical history variable (pneumonia; temperature spikes; aspi ration pneumonia; intubation/tracheostomy), behavioural variables (alertness; cooperativeness/agitation; attention/ interaction ability; awareness of swallowing problem; ability to manage secretions), gross motor function (postural control; fatigability), oral motor test results (oral, pharyngeal, lar yngeal anatomy and physiology; ability to follow directions; dysarthria; facial weakness; oral apraxia; oral sensation; pharyngeal wall contraction on gag; saliva swallowing; voluntary cough, throat clearing), observation during trial swallows of 1 cc thin liquid, 1 cc pudding, ¼ cookie (apraxia of swallowing; oral residue, coughing/throat clearing; delayed pharyngeal swallow; reduced laryngeal elevation; gurgly voice; multiple swallows per bolus), total number of unsafe observations, total number of unsafe observations on behavioural and gross motor function, total number of unsafe observations made during oral motor testing and observations during trial swallows Endpoint: several endpoints are described
Mann 2002	Patients admitted to a hospital with a first acute stroke. Thirty per cent were between 0 and 64 years; 39% between 65 and 74 years; 22% between 75 and 84 years; 9% ≥85 years	128	Type of test: VF Swallow material: two or three times, in upright position, 5 and 10 mL of thin liquid, thick li quid and semisolid barium preparations, consecutively Endpoint: aspiration and dysphagia	Type of test: clinical assessment including oral-motor-sensory examination (voice, speech, and language function) and observation of swallowing saliva initially and subsequently 5 mL of water, 20 mL of water and a thickened fluid if appropriate Endpoint: cut-off score on MASA for severity of dysphagia and aspiration

Table 4 (*Continued*)

References	Population	*n*	Reference test	Index test
Mari *et al.* 1997	Patients (inpatients and outpatients) admitted to a rehabilitation clinic with a neurological disorder (acute stroke, Parkinson's disease, traumatic brain injury, multiple sclerosis, amyotrophic lateral sclerosis, myotonic dystrophy, abiotrophic disease. Mean age 59·8 years (18–80)	93	Type of test: VF Swallow material: 5 or 10 or 15 mL fluid and solid barium boluses Endpoint: aspiration	Type of test: clinical assessment: a 25-item form and a 3 oz water swallow test Swallow material: 9 mL water to drink from a cup without interruption Endpoint: coughing or wet or hoarse voice quality
McCullough *et al.* 2001	Patients with recent stroke. Mean age 67·8 years (40–96)	60	Type of test: VF Swallow material: two swallows of thin liquid (5 mL) followed by two 10 cc swallows of thin liquid. Next two 5 cc boluses of thick liquid followed by two 10 cc boluses of thick liquid for lateral view. Then two more 5 cc thick li quid for a posterior-anterior view In addition, again in lateral position, two swal lows of 5 cc applesauce from a spoon followed by two solids. All test material was mixed/ impregnated with barium Endpoint: aspiration	Type of test: clinical bedside examination, history, oral-motor (speech and praxis), voice, and trial swallows (thin liquid and thick liquid from a pill cup, puree, and solid from a spoon, administered in 5 cc boluses (1/4 cookie for solid) Endpoint: delayed swallow, total swallow duration, laryngeal elevation, and swallows per bolus, rated using Logeman's four finger method. These were scored binarily and raters were not trained
Smith *et al.* 2000	Patients with acute stroke admitted to a hospital. Median age 69 years (51–90)	53	Type of test: VF Swallow material: various amounts and consis tencies of barium sulphate, starting with 3 mL and gradually increasing aliquots. These comprised 5, 10 and 20 mL of thick liquid barium, the same quantities of dilute liquid barium, then 5 mL yoghurt and finally 5 mL solid (bread plus barium) Endpoint: aspiration and/or penetration	Type of test: bedside swallowing assessment and pulse oximetry Swallow material: variety of quantities and consistencies Endpoint: subjective assessment of aspiration and a desaturation of ≥2%
Smithard *et al.* 1998	Patients within 24 hours of the onset of acute stroke. Median age 79 years (40–93)	94	Type of test: VF Swallow material: different consistencies and volumes of barium Endpoint: aspiration	Type of test: medical bedside assessment: first 5 mL spoonful of water, then swallowing 60 mL of water within 2 minutes. Coughing and/or choking on more than one occasion of three attempts and/or the presence of a wet voice was considered unsafe SLT bedside assessment: clinical judgment as to whether a swallow was safe or unsafe Endpoint: see above

Table 4 (*Continued*)

References	Population	*n*	Reference test	Index test
Trapl *et al.* 2007	Patients with first ever acute stroke. Mean age group 1:74·6 (SE 2·4); mean age group 2:76·8 (SE 1·85)	50	Type of test: FEES Swallow material: no information available Endpoint: score between 4 and 5 of the Penetration Aspiration Scale of Rosenbek indicating laryngeal penetration of swallow material reaching to the vocal folds	Type of test: Gugging Swallowing Screen (GUSS) consisting of two parts: one indirect swallowing test consisting of a simple saliva test; two direct swallowing test consisting of three sequentially performed subsets, starting with semisolid, then liquid, and finally soli textures Endpoint: between 14 and points on GUSS score indicating risk of aspiration

VF, videofluroscopic swallowing examination; FEES, fibreoptic endoscopic evaluation of swallowing; SLP, speech language pathologist; SLT, speech language therapist.

were only achieved using a subjective assessment of the risk of aspiration, as was done by McCullough *et al.* (2001). Smith *et al.* (2000) also used a subjective assessment of aspiration and, like McCullough *et al.*, found sufficient sensitivity and specificity. The positive likelihood ratios ranged from 1·3 to 3·7, and the negative likelihood ratios from 0·0 to 0·8.

Oxygen desaturation

Three studies used oxygen desaturation to evaluate aspiration/penetration (Smith *et al.* 2000, Lim *et al.* 2001, Chong *et al.* 2003). They defined a desaturation of ≥2% as the endpoint: sensitivity ranged from 56% to 87%, specificity from 39% to 97% (Table 5). Positive likelihood ratios ranged from 1·4 to 18·9 and negative likelihood ratios from 0·3 to 0·5. Only the study by Lim *et al.* (2001) achieved our predefined values for sensitivity and specificity.

Swallow test in combination with oxygen desaturation

Three studies involved a combination of a swallow test and oxygen desaturation (Smith *et al.* 2000, Lim *et al.* 2001, Chong *et al.* 2003). Two tests used coughing, choking, voice change or ≥2% desaturation as the endpoints (Lim *et al.* 2001, Chong *et al.* 2003). The sensitivities of these two tests were 94% and 98%, respectively, while the specificities were 63% and 70%, respectively. Positive likelihood ratios were 2·5 and 3·3, while negative likelihood ratios were 0·0 and 0·1. Smith *et al.* (2000) used a combination of subjective assessment of aspiration and ≥2% oxygen desaturation as the endpoint. It found a sensitivity of 73%, a specificity of 76%, a positive likelihood ratio of 3·1, and a negative likelihood ratio of 0·3.

Clinical features

Three researchers reported on a variety of clinical features used to detect aspiration and/or penetration (Daniels *et al.* 1997, Logemann *et al.* 1999, McCullough *et al.* 2001). The clinical features generally had either low sensitivity or low specificity, or both. Positive likelihood ratios ranged from 1·1 to 3·2, and negative likelihood ratios from 0·1 to 1·0. Only the study by Daniels *et al.* (1997) showed a sensitivity of 73% and a specificity of 72% for dysphonia, which are satisfactory values. The positive and negative likelihood ratios were 2·7 and 0·4, respectively.

History components

One report described the use of a history of poor nutrition to assess aspiration (McCullough *et al.* 2001). Sensitivity and specificity were 50% and 76%, respectively, while positive and negative likelihood ratios were 2·1 and 0·7, respectively.

Table 5 Diagnostic performance of the various tests in studies with sufficient methodological quality

Bedside test	Endpoint of reference test	Endpoint of index test	Prevalence		Sensitivity		Specificity		LR+	LR−
			%	(95% CI)	%	(95% CI)	%	(95% CI)	(95% CI)	(95% CI)
Water test										
Chong *et al.* 2003	Aspiration or penetration	Coughing, choking or voice change	68	53–80	79	62–91	63	36–84	2·1 (1·1–4·1)	0·3 (0·2–0·7)
Daniels *et al.* 1997	Aspiration or penetration	Coughing	44	31–57	65	44–82	79	61–90	3·1 (1·5–6·3)	0·4 (0·3–0·8)
		Voice change	44	31–57	27	12–48	88	71–96	2·2 (0·7–6·8)	0·8 (0·7–1·1)
Lim *et al.* 2001	Aspiration	Coughing, choking or voice change	52	38–66	85	64–95	75	53–89	3·4 (1·7–6·9)	0·2 (0·1–0·5)
Mari *et al.* 1997	Aspiration	Coughing	46	36–57	51	36–66	86	73–94	3·6 (1·7–7·7)	0·6 (0·4–0·8)
Smithard *et al.* 1998	Aspiration	Coughing, choking, wet voice assessed by doctor	21	14–31	70	46–87	66	54–77	2·1 (1·3–3·2)	0·5 (0·2–0·9)
		Coughing, choking, wet voice assessed by SLT	23	15–34	47	25–71	86	74–93	3·4 (1·6–7·3)	0·6 (0·4–0·9)
Trial swallow										
Logemann *et al.* 1999	Aspiration	Cough or throat clear	54	46–61	78	68–85	58	47–68	1·8 (1·4–2·4)	0·4 (0·3–0·6)
		Gurgly voice	54	46–61	41	32–51	76	66–84	1·7 (1·1–2·7)	0·8 (0·7–0·9)
		Reduced laryngeal elevation	54	46–61	66	56–75	57	46–67	1·5 (1·2–2·0)	0·6 (0·4–0·8)
		Multiple swallows	54	46–61	58	48–67	57	46–67	1·3 (1·0–1·8)	0·7 (0·6–0·9)
McCullough *et al.* 2001	Aspiration	Spontaneous cough	37	25–50	68	45–85	82	65–92	3·7 (1·8–7·7)	0·4 (0·2–0·7)
		Wet voice	37	25–50	50	29–71	63	46–78	1·4 (0·8–2·4)	0·8 (0·5–1·2)
		Subjective estimate of aspiration	37	25–50	78	54–91	63	46–78	2·1 (1·3–3·4)	0·4 (0·2–0·8)
Smith *et al.* 2000	Aspiration or penetration	Subjective assessment of aspiration	28	17–43	80	51–95	68	51–82	2·5 (1·5–4·3)	0·3 (0·1–0·8)
Trapl *et al.* 2007	Laryngeal penetration	Risk of aspiration on GUSS	55	40–69	100	84–100	63	41–82	2·8 (1·6–4·8)	0 (0–un-defined)
Oxygen desaturation										
Chong *et al.* 2003	Aspiration or penetration	≥2% desaturation	67	53–79	56	38–72	97	73–1·0	18·9 (1·2–295·3)	0·5 (0·3–0·7)
Lim *et al.* 2001	Aspiration	≥2% desaturation	52	38–66	77	56–90	83	62–95	4·6 (1·8–11·6)	0·3 (0·1–0·6)
Smith *et al.* 2000	Aspiration or penetration	≥2% desaturation	28	17–43	87	58–98	39	24–57	1·4 (1·0–1·9)	0·3 (0·1–1·3)
Swallow test + oxygen desaturation										
Chong *et al.* 2003	Aspiration or penetration	Coughing, choking or voice change or ≥2% desaturation	68	53–80	94	79–99	63	36–84	2·5 (1·3–4·6)	0·1 0·0–0·4
Lim *et al.* 2001	Aspiration	Coughing, choking or voice change or ≥2% desaturation	52	38–66	98	82–100	70	49–86	3·3 (1·8–6·0)	0·0 0·0–0·4
Smith *et al.* 2000	Aspiration or penetration	Subjective assessment of aspiration + ≥ 2% desaturation	28	17–43	73	45–91	76	59–88	3·1 (1·6–5·9)	0·3 0·1–0·8

Table 6 Diagnostic performance of clinical features from studies with sufficient methodological quality

Clinical features	Endpoint reference test	Feature/endpoint†	Prevalence		Sensitivity		Specificity		LR+ (95% CI)	LR− (95% CI)
			%	95% CI	%	95% CI	%	95% CI		
Daniels *et al.* 1997	Aspiration or penetration	Dysphonia	44	31–57	73	52–88	72	54–86	2·7 (1·5–4·9)	0·4 (0·2–0·7)
McCullough *et al.* 2001*	Aspiration	Dysphonia	37	25–50	98	79–100	27	15–44	1·3(1·1–1·6)	0·1 (0·0–1·4)
Daniels *et al.* 1997	Aspiration or penetration	Dysarthria	44	31–57	85	64–95	55	37–71	1·9 (1·2–2·8)	0·3 (0·1–0·7)
McCullough *et al.* 2001	Aspiration	Dysarthria	37	25–50	77	54–91	55	38–71	1·7 (1·1–2·6)	0·4 (0·2–0·9)
Daniels *et al.* 1997	Aspiration or penetration	Abnormal gag	44	31–57	54	34–73	67	48–81	1·6 (0·9–2·9)	0·7 (0·4–1·1)
Logemann *et al.* 1999	Aspiration	Abnormal gag	54	46–61	33	24–43	81	71–88	1·7 (1·0–2·8)	0·8 (0·7–1·0)
Daniels *et al.* 1997	Aspiration or penetration	Volitional cough	44	31–57	42	24–63	85	67–94	2·8 (1·1–7·0)	0·7 (0·5–1·0)
McCullough *et al.* 2001	Aspiration	Volitional cough/quality	37	25–50	68	45–85	45	30–62	1·2 (0·8–1·8)	0·7 (0·4–1·4)
McCullough *et al.* 2001	Aspiration	Management of secretions	37	25–50	50	29–71	84	68–93	3·2 (1·4–7·4)	0·6 (0·4–0·9)
McCullough *et al.* 2001	Aspiration	Intelligibility	37	25–50	73	50–88	58	41–73	1·7 (1·1–2·7)	0·5 (0·2–1·0)
McCullough *et al.* 2001	Aspiration	Tongue strength	37	25–50	50	29–71	74	57–86	1·9 (1·0–3·7)	0·7 (0·4–1·0)
McCullough *et al.* 2001	Aspiration	Wet voice (ah)	37	25–50	50	29–71	84	68–93	3·2 (1·4–7·4)	0·6 (0·4–0·9)
McCullough *et al.* 2001	Aspiration	Wet voice (speech)	37	25–50	50	29–71	79	62–90	2·4 (1·1–5·0)	0·6 (0·4–1·0)
McCullough *et al.* 2001	Aspiration	Voice strained	37	25–50	50	29–71	74	57–86	1·9 (1·0–3·7)	0·7 (0·4–1·0)
McCullough *et al.* 2001	Aspiration	Palatal symmetry	37	25–50	50	29–71	53	36–69	1·1 (0·6–1·8)	1·0 (0·6–1·5)
History components										
McCullough *et al.* 2001	Aspiration	Poor nutrition	37	25–50	50	29–71	76	59–88	2·1 (1·0–4·3)	0·7 (0·4–1·0)
Standardized form with clinical features										
Leder & Espinosa 2002	Aspiration risk	≥2 variables present	45	31–60	86	64–96	30	14–50	1·2 (0·9–1·6)	0·5 (0·1–1·6)
Logemann *et al.* 1999	Aspiration	≥5 unsafe ratings	54	46–61	58	48–67	60	50–70	1·5 (1·1–2·0)	0·7 (0·6–0·9)
Mann 2002	Dysphagia (mild, moderate, severe)	Dysphagia (definite/probable/possible)	64	55–72	73	62–82	89	76–96	6·7 (2·9–15·6)	0·3 (0·2–0·4)
	Aspiration (mild, moderate, severe)	Aspiration (definite/probable/possible)	22	15–30	93	75–99	63	53–72	2·5 (1·9–3·3)	0·1 (0·0–0·4)

*Only clinical features of their study with a sensitivity and specificity ≥50% are included in this table.
†The endpoints of the clinical features were all rated binarily (yes/no; present/absent; normal/abnormal).

Standardized forms

Three papers described the use of a standardized form with various clinical identifiers to assess aspiration (Logemann *et al.* 1999, Leder & Espinosa 2002, Mann 2002). All three had sufficient methodological quality. Sensitivity ranged from 58% to 93%, and specificity from 30% to 63%. The positive likelihood ratios ranged from 1·2 to 6·7, the negative likelihood ratios from 0·1 to 0·7. Only the Mann Assessment of Swallowing Ability (MASA) (Mann 2002) met our predetermined psychometric properties.

Feasibility

Feasibility was assessed in terms of the time required to conduct the screening test and the complexity of the test in terms of the number of test materials needed. Of the seven studies (Smithard *et al.* 1998, Smith *et al.* 2000, Lim *et al.* 2001, McCullough *et al.* 2001, Mann 2002, Chong *et al.* 2003, Trapl *et al.* 2007) that met the predetermined criteria for sensitivity and specificity, only the report by Mann (2002) said that the examination could be administered in about 15–20 minutes. None of the other authors mentioned the time required to administer the screening test. Data on test materials required were derived from the descriptions of the screening procedure: water tests required fewer materials than all of the other trial swallow tests.

Discussion

In this review, we determined the effectiveness and feasibility of bedside screening methods that can be used by nurses with patients with neurological disorders for detecting dysphagia. Although dysphagia needs a multidisciplinary approach, we think that nurses, if well trained, have an important role to play in screening and observing, because of their 24-hour availability. International guidelines also recommend screening by appropriately trained clinicians, i.e. nurses or other staff (Royal College of Physicians 2004, Canadian Stroke Network and the Heart and Stroke Foundation of Canada: Canadian Stroke Strategy 2006, Kwaliteitsinstituut voor de Gezondheidszorg: CBO 2008). Perry (2001) showed that trained nurses were competent in performing a standardized swallowing screening. However, patients who fail such a screening should be referred as soon as possible for more detailed assessment by speech and language therapists so that further strategies can be designed.

Our search resulted in an abundance of studies involving different screening methods, different applications of reference tests and different cut-off points for reference tests and index tests. There was great variety in methodological quality, with a majority of the studies (70%) showing major methodological flaws. In order to produce a reliable systematic review, we used only those studies with the highest quality rating to calculate psychometric properties, as suggested by Deeks (2001). Based on these psychometric properties, we concludes that a water test combined with pulse oximetry, using several endpoints (choking, coughing and voice change) offers the most promising results as a screening tool.

Although pulse oximetry for screening has been criticized in the literature (Collins & Bakheit 1997, Colodney 2001), it has also been shown to detect silent aspiration (Lim *et al.* 2001, Chong *et al.* 2003), which cannot be detected by a swallow test alone. Silent aspiration is the aspiration of swallowed material without any clinical manifestations of aspiration such as coughing and may be prevalent in 15–39% of patients with subacute stroke and in 2–25% of unselected patients with acute stroke (Ramsey *et al.* 2005). A review by Ramsey *et al.* (2005) produced conflicting results on the consequences of silent aspiration: some studies suggested increased frequency of chest infections and poorer clinical outcomes, while others found no adverse effects in healthy individuals. More research is needed on the consequences of silent aspiration and on the value of oxygen desaturation in addition to a trial swallow.

Four studies assessed the value of administering various liquids, semi-solids and solids to assess risk of aspiration. The Gugging Swallowing Screen (Trapl *et al.* 2007) and studies which examined subjective assessment of aspiration found satisfactory psychometric properties (Smith *et al.* 2000, McCullough *et al.* 2001, Trapl *et al.* 2007). The advantage of evaluating patients' ability to swallow materials of different consistencies is that it approximates their normal daily food habits. However, it requires considerable quantities of test materials, such as different liquids, semi-solids and solids, as well as equipment to administer these, making it less feasible than a test based on water alone. The same argument applies to the use of the MASA; although Mann (2002) found high sensitivity and specificity values for this test, and stated that the examination can be administered in 15–20 minutes, it is less practical because of the large number of items and the need to calculate cut-off scores for either dysphagia or aspiration, and the same goes for the Gugging Swallowing Screen. The availability of a simpler test with good psychometric properties reduces the need to use more complex tests for screening in practice.

Our review also revealed that single clinical features, such as abnormal gag or volitional cough on its own or medical

history components, are not very useful for identifying patients at risk of aspiration. The sensitivity and specificity of these features were generally low.

We were only able to compare the selected tests in a descriptive manner rather than by statistical pooling, as they differed greatly in the methods used to apply similar bedside tests, for instance in the way in which water or other test materials were administered. The protocols used for the reference tests also varied widely. In addition, there was considerable variety in the choice of endpoints, either in the reference test or the index tests. Some authors used aspiration for the reference test, while others used aspiration and/or penetration. The same problem arose for the index test: some authors presented results for each endpoint, while others looked at a combination of features as the endpoint. All of this hampers a comparison of the bedside tests examined in the various studies and makes it impossible to pool data. In such circumstances, some authors have recommended using the diagnostic odds ratio, which can be calculated as positive likelihood ratio/negative likelihood ratio (Glas *et al.* 2003). This test statistic ranges from zero to infinity, with higher values indicating better test performance, and may be a better measure of a test's discriminatory performance than sensitivity and specificity values. However, since it does not distinguish between the two types of diagnostic errors, it is only useful when the balance between false-negative and false-positive rates is not of immediate importance (Glas *et al.* 2003). In terms of screening for dysphagia, our intention was to detect patients who aspirate, and so high sensitivity was chosen as a criterion for selecting screening methods, because missing a patient may result in serious adverse events (Smithard *et al.* 1996, Hinchey *et al.* 2005). At the same time, we regarded it as acceptable that some patients might unnecessarily undergo further diagnostic examination, although we set the lower limit for specificity at 60% because of the extra healthcare costs that result from unnecessary diagnostic procedures.

Although this systematic review focused on bedside screening to detect dysphagia in patients with neurological disorders, most bedside tests found in the literature were primarily used with patients with stroke. Some authors included patients with neurological disorders as well as those with other medical conditions. In our quality rating, these studies were allocated a plus/minus for item 7 of our quality rating list, which refers to generalization. De Vet *et al.* (2001) suggested that researchers should present diagnostic accuracy for various subgroups of populations to facilitate the extrapolation of the results to the population of interest. As this was not done in most of the studies included in our review, caution should be exercised in generalizing the results of our review to other neurological conditions. Further research on dysphagic patients with other neurological conditions is necessary, and should not be restricted to the validation of screening tools, but should also address the effects of screening.

Videofluoroscopy and FEES were used as the gold standard for the screening test, because both are considered equally valuable for the detection of dysphagia (Langmore *et al.* 1991, Doggett *et al.* 2002, Rao *et al.* 2003). However, variations have been reported between assessors in rating the videos from both instrumental examinations (Wu *et al.* 1997). The highest levels of agreement and reliability are reached when the final score is based on consensus between expert raters who are experienced in interpreting videos (Scott *et al.* 1998). In the current review, most papers did not clearly describe how the final score on the reference test had been obtained. If it was based on a single judgment, the conclusions on the psychometric properties of the bedside test may have been biased.

To our best knowledge, this is the first systematic review to summarize properties of bedside tests in patients with neurological disorders. Other reviews were limited to patients with stroke (Martino *et al.* 2000, Perry & Love 2001, Kalf 2002, Ramsey *et al.* 2003, Westergren 2006), were not systematic (Lambert & Gisel 1996), or did not assess the risk of bias (Lambert & Gisel 1996, Perry & Love 2001, Ramsey *et al.* 2003). In line with our findings, all other reviewers recognized that the variety of tests made comparisons very difficult. The only reviews to have cautious conclusions were those by Martino *et al.* (2000) (50 mL water and impaired pharyngeal sensation) and Kalf (2002) (water combined with pulse oximetry), with the results of the latter being in line with our findings and those recommended in the new Dutch stroke guidelines (Kwaliteitsinstituut voor de Gezondheidszorg: CBO 2008). Although other international guidelines stress the importance of screening, they do not recommend a specific screening method (Royal College of Physicians 2004; Canadian Stroke Network and the Heart and Stroke Foundation of Canada: Canadian Stroke Strategy 2006). Martino *et al.* (2000) and Kalf (2002) both assessed the methodological quality of the studies using clearly specified criteria. However, Martino's conclusion was mainly based on the study by Kidd *et al.* (1993), which was assessed in our review as being of doubtful methodology. Although Martino *et al.* did assess the quality of their included studies, they disregarded the methodological scores in their conclusions. In addition, all of these reviews were published more than 6 years ago and may be out of date. Only Westergren (2006) included studies up to 2004, but his review was limited to patients with stroke and used only a very crude estimation to

What is already known about this topic

- Dysphagia is very common in patients with neurological disorders such as stroke, Parkinson's disease and multiple sclerosis.
- It is important to screen patients as soon as possible after dysphagia is suspected to minimize risks to health.
- Many bedside tests screening for dysphagia are available.

What this paper adds

- A water test combined with pulse oximetry using coughing, choking and voice alteration as the endpoints is currently the best method to screen patients for dysphagia.
- Single clinical features such as abnormal gag, volitional cough or medical history components are not useful to identify patients at risk of aspiration/penetration.
- Further research is needed to determine the most effective procedure for administrating water tests and the value of pulse oximetry in addition to a trial swallow.

Implications for practice and/or policy

- Standardized protocols for instrumental testing and bedside screening should be developed and used.
- A water test combined with pulse oximetry should be used by nurses (or other staff) for screening patients with neurological disorders at risk of dysphagia.

assess the methodological shortcomings of the included studies.

It is well known that systematic reviews are prone to selection bias, especially if they review diagnostic studies, which means that we may not have included all published studies on screening. Search strategies for papers reporting on diagnostic studies have been less extensively researched than those in the domain of clinical trials. However, we used a very extensive search strategy, recommended by the Dutch Cochrane Collaboration and based on the study by Deville *et al.* (2000), whose strategy resulted in a sensitivity close to 90%. However, for practical reasons we focused on studies in English, Dutch and German. We know from the literature that studies with statistically significant results are more likely to be published in English (Egger *et al.* 1997, Egger 1998), leading to publication bias. Although it is difficult to examine the effect of such publication bias in our review, we do need to take into account the possibility that the included studies were those with the most positive results.

Conclusion

Based on a combination of the best psychometric properties and the feasibility of the screening methods, a water test using a mix of endpoints (coughing, choking and voice alteration) combined with pulse oximetry (desaturation ≥2%) produces the best results in terms of detecting patients with dysphagia in practice. However, performing this screening requires training.

Unfortunately, the literature does not give clear indications about the way in which the water test should be administered. More research is needed to achieve further improvements to the water test and to assess the value of pulse oximetry in combination with a trial swallow, especially for the detection of silent aspiration. Finally, we recommend developing and using standardized protocols for instrumental testing as well as bedside screening, in order to allow comparisons of the results of future studies.

Funding

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Author contributions

GB was responsible for the study conception and design. GB and RS performed the data collection. GB and RS performed the data analysis. GB was responsible for the drafting of the manuscript. JL, ML and RdW made critical revisions to the paper for important intellectual content.

References

Addington W.R., Stephens R.E., Gilliland K. & Rodriguez M. (1999) Assessing the laryngeal cough reflex and the risk of developing pneumonia after stroke. *Archives of Physical Medicine and Rehabilitation* **80**(2), 150–154.

Agency for Health Care Policy and Research. (1999) *Diagnosis and Treatment of Swallowing Disorders (dysphagia) in Acute-Care Stroke Patients*. U.S. Department of Health and Human Services, AHCPR, Rockville, MD.

Canadian Stroke Network and the Heart and Stroke Foundation of Canada: Canadian Stroke Strategy. (2006) *Canadian Best Practice Recommendations for Stroke Care: 2006*. Canadian Stroke Network and the Heart and Stroke Foundation of Canada: Canadian Stroke Strategy, Ottawa.

Chong M.S., Lieu P.K., Sitoh Y.Y. & Meng Y.Y. (2003) Bedside clinical methods useful as screening test for aspiration in elderly

patients with recent and previous strokes. *Annals Academy of Medicine Singapore* **32**(6), 790–794.

Cochrane Collaboration. (2008) *Cochrane Handbook for Diagnostic Test Accuracy*. Retrieved from http://srdta.cochrane.org/en/-authors.html.

Collins M.J. & Bakheit A.M. (1997) Does pulse oximetry reliably detect aspiration in dysphagic stroke patients? *Stroke* **28**(9), 1773–1775.

Colodney N. (2001) Effects of age, gender, disease, and multisystem involvement on oxygen saturation levels in dysphagic persons. *Dysphagia* **16**, 48–57.

Daniels S.K., McAdam C.P., Brailey K. & Foundas A.L. (1997) Clinical assessment of swallowing and prediction of dysphagia severity. *American Journal of Speech Language Pathology* **6**(4), 17–24.

Daniels S.K., Brailey K., Priestly D.H., Herrington L.R., Weisberg L.A. & Foundas A.L. (1998) Aspiration in patients with acute stroke. *Archives of Physical Medicine and Rehabilitation* **79**(1), 14–19.

De Vet H.C.W., van der Weijden T., Muris J.W.M., Heyrman J., Buntinx F. & Knottnerus J.A. (2001) Systematic reviews of diagnostic research. Considerations about assessment and incorporation of methodological quality. *European Journal of Epidemiology* **17**, 301–306.

Deeks J.J. (2001) Systematic reviews of evaluations of diagnostic and screening tests. *British Medical Journal* **323**, 157–162.

DePippo K.L., Holas M.A. & Reding M.J. (1992) Validation of the 3-oz water swallow test for aspiration following stroke. *Archives of Neurology* **49**(12), 1259–1261.

Deville W.L.J.M., Bezemer P.D. & Bouter L.M. (2000) Publications on diagnostic test evaluation in family medicine journals: an optimal search strategy. *Journal of Clinical Epidemiology* **53**, 65–69.

Doggett D.L., Turkelson C.M. & Coates V. (2002) Recent developments in diagnosis and intervention for aspiration and dysphagia in stroke and other neuromuscular disorders. *Current Atherosclerosis Reports* **4**(4), 311–318.

Egger M. (1998) Meta-analysis bias in location and selection of studies. *British Medical Journal* **316**, 61–66.

Egger M., Zellweger-Zähner T., Schneider M., Junker C., Lengeler C. & Antes G. (1997) Language bias in randomized controlled trials published in English and German. *The Lancet* **350**, 326–329.

Garon B.R., Engle M. & Ormiston C. (1995) Reliability of the 3-oz water swallow test utilizing cough reflex as sole indicator of aspiration. *Journal of Neurologic Rehabilitation* **9**(3), 139–143.

Glas A.F., Lijmer J.G., Prins M.H., Nonsel G.J. & Bossuyt P.M.M. (2003) The diagnostic odds ratio: a single indicator of test performance. *Journal of Clinical Epidemiology* **56**, 1129–1135.

Higo R., Tayama N., Watanabe T. & Nito T. (2003) Pulse oximetry monitoring for the evaluation of swallowing function. *European Archives of Oto-rhino-pharyngology* **260**(3), 124–127.

Hinchey J.A., Shepherd T., Furie K., Smith D., Wang D. & Tonn S. (2005) Formal dysphagia screening protocols prevent pneumonia. *Stroke* **36**, 1972–1976.

Horner J. & Massey E.W. (1988) Silent aspiration following stroke. *Neurology* **38**(2), 317–319.

Horner J., Massey E.W., Riski J.E., Lathrop D.L. & Chase K.N. (1988) Aspiration following stroke: clinical correlates and outcome. *Neurology* **38**(9), 1359–1362.

Horner J., Brazer S.R. & Massey E.W. (1993) Aspiration in bilateral stroke patients: a validation study. *Neurology* **43**(2), 430–433.

Kalf J.G. (2002) Slikscreening na een beroerte: een evidence based review. *Logopedie en Foniatrie* **74**(3), 76–83.

Kidd D., Lawson J., Nesbitt R. & MacMahon J. (1993) Aspiration in acute stroke: a clinical study with videofluoroscopy. *Quarterly Journal of Medicine* **86**(12), 825–829.

Koopman W.J., Wiebe S., Colton-Hudson A., Moosa T., Smith D. & Nicolle M.W. (2004) Prediction of aspiration in myasthenia gravis. *Muscle and Nerve* **29**(2), 256–260.

Kwaliteitsinstituut voor de Gezondheidszorg: CBO. (2008) *Diagnostiek, behandeling en zorg voor patiënten met een beroerte*. CBO, Utrecht.

Lam K., Lam F.K., Lau K.K., Chan Y.K., Kan E.Y., Woo J., Wong F.K. & Ko A. (2007) Simple clinical tests may predict severe oropharyngeal dysphagia in Parkinson's disease. *Movement Disorders* **22**(5), 640–644.

Lambert H.C. & Gisel E.G. (1996) The assessment of oral, pharyngeal and esophageal dysphagia in elderly persons. *Physical & Occupational Therapy in Geriatrics* **14**(4), 1–25.

Langmore S.E., Schatz K. & Olson N. (1991) Endoscopic and videofluoroscopic evaluations of swallowing and aspiration. *Annals of Otology, Rhinology, and Laryngology* **100**(8), 678–681.

Leder S.B. & Espinosa J.F. (2002) Aspiration risk after acute stroke: comparison of clinical examination and fiberoptic endoscopic evaluation of swallowing. *Dysphagia* **17**(3), 214–218.

Lijmer J.G., Mol B.W., Heisterkamp S., Bonsel G.J., Prins M.H., van der Meulen J.H.P. & Bossuyt P.M.M. (1999) Empiracel evidence of design-related bias in studies of diagnostic tests. *Journal of the American Medical Association* **282**(11), 1061–1063.

Lim S.H.B., Lieu P.K., Phua S.Y., Seshadri R., Venketasubramanian N., Lee S.H. & Choo P.W.J. (2001) Accuracy of bedside clinical methods compared with fiberoptic endoscopic examination of swallowing (FEES) in determining the risk of aspiration in acute stroke patients. *Dysphagia* **16**(1), 1–6.

Linden P. & Siebens A.A. (1983) Dysphagia: predicting laryngeal penetration. *Archives of Physical Medicine and Rehabilitation* **64**(6), 281–284.

Logemann J.A., Veis S. & Colangelo L. (1999) A screening procedure for oropharyngeal dysphagia. *Dysphagia* **14**(1), 44–51.

Mann G.D. (2002) MASA: the Mann Assessment of Swallowing Ability. In *Dysphagia Series* (the author, ed.), Singular Thomson Learning, New York, NY, pp. 56.

Mann G., Hankey G.J. & Cameron D. (1999) Swallowing function after stroke: prognosis and prognostic factors at 6 months. *Stroke* **30**(4), 744–748.

Mari F., Matei M., Ceravolo M.G., Pisani A., Montesi A. & Provinciali L. (1997) Predictive value of clinical indices in detecting aspiration in patients with neurological disorders. *Journal of Neurology, Neurosurgery, and Psychiatry* **63**(4), 456–460.

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

Massey R. & Jedlicka D. (2002) The Massey bedside swallowing screen. *Journal of Neuroscience Nursing* **34**(5), 252–253, 257–260.

McCullough G.H., Wertz R.T. & Rosenbek J.C. (2001) Sensitivity and specificity of clinical/bedside examination signs for detecting

aspiration in adults subsequent to stroke. *Journal of Communication Disorders* **34**(1–2), 55–72.

Murray J. (ed.) (1999a) The clinical swallowing examination. In *Manual of Dysphagia Assessment in Adults*, Singular, San Diego, CA, pp. 37–112.

Murray J. (ed.) (1999b) *Manual of Dysphagia Assessment in Adults*. Singular, San Diego, CA.

NANDA. (2007) *Nursing Diagnosis: Definitions & Classifications 2007–2008*. NANDA International, Philadelphia, PA.

Nishiwaki K., Tsuji T., Liu M., Hase K., Tanaka N. & Fujiwara T. (2005) Identification of a simple screening tool for dysphagia in patients with stroke using factor analysis of multiple dysphagia variables. *Journal of Rehabilitation Medicine* **37**(4), 247–251.

Ott D.J. & Pikna L.A. (1993) Clinical and videofluoroscopic evaluation of swallowing disorders. *American Journal of Roentgenology* **161**(3), 507–513.

Perry L. (2001) Screening swallowing function of patient with acute stroke. Part two: detailed evaluation of the tool used by nurses. *Journal of Clinical Nursing* **10**(4), 474–481.

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

Prosiegel M., Schelling A. & Wagner-Sonntag E. (2004) Dysphagia and multiple sclerosis. *The International MS Journal* **11**(1), 22–31.

Ramsey D.J., Smithard D.G. & Kalra L. (2003) Early assessments of dysphagia and aspiration risk in acute stroke patients. *Stroke* **34**(5), 1252–1257.

Ramsey D., Smithard D. & Kalra L. (2005) Silent aspiration: what do we know? *Dysphagia* **20**, 218–225.

Rao N., Brady S.L., Chaudhuri G., Donzelli J.J. & Wesling M.W. (2003) Gold-standard?: analysis of the videofluoroscopic and fiberoptic endoscopic swallow examinations. *Journal of Applied Research* **3**(1), 89–96.

Rosenbek J.C., McCullough G.H. & Wertz R.T. (2004) Is the information about a test important?: applying the methods of evidence-based medicine to the clinical examination of swallowing. *Journal of Communication Disorders* **37**(5), 437–450.

Royal College of Physicians. (2004) *National Clinical Guidelines for Stroke*. Prepared by the Intercollegiate Stroke Working Party, RCP, London.

Scott A., Perry A. & Bench J. (1998) A study of interrater reliability when using videofluoroscopy as an assessment of swallowing. *Dysphagia* **13**(4), 223–227.

Shaw J.L., Sharpe S., Dyson S.E., Pownall S., Wa;ters S., Saul C., Enderby P., Healy K. & O'Sullivan H. (2004) Bronchial auscultation: an effective adjunct to speech and language therapy bedside assessment when detecting dysphagia and aspiration? *Dysphagia* **19**(4), 211–218.

Sherman B., Nisenboum J.M., Jesberger B.L., Morrow C.A. & Jesberger J.A. (1999) Assessment of dysphagia with the use of pulse oximetry. *Dysphagia* **14**(3), 152–156.

Smith Hammond C.A. & Goldstein L.B. (2006) Cough and aspiration of food and liquids due to oral-pharyngeal dysphagia: ACCP evidence-based clinical practice guidelines. *Chest* **129**(Suppl. 1), 154S–168S.

Smith H.A., Lee S.H., O'Neill P.A. & Connolly M.J. (2000) The combination of bedside swallowing assessment and oxygen saturation monitoring of swallowing in acute stroke: a safe and humane screening tool. *Age and Ageing* **29**(6), 495–499.

Smithard D.G., O'Neill P.A., Park C., Morris J., Wyatt R., England R. & Martin D.F. (1996) Complications and outcome after acute stroke: does dysphagia matter? *Stroke* **27**(7), 1200–1204.

Smithard D.G., O'Neill P.A., Park C., England R., Renwick D.S., Wyatt R., Morris J. & Martin D.F. (1998) Can bedside assessment reliably exclude aspiration following acute stroke? *Age and Ageing* **27**(2), 99–106.

Splaingard M.L., Hutchins B., Sulton L.D. & Chaudhuri G. (1988) Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment. *Archives of Physical Medicine and Rehabilitation* **69**(8), 637–640.

Tohara H., Saitoh E., Mays K.A., Kuhlemeier K.V. & Palmer J.B. (2003) Three tests for predicting aspiration without videofluorography. *Dysphagia* **18**(2), 126–134.

Trapl M., Enderle P., Nowotny M., Teuschl Y., Matz K., Dachenhausen A. & Brainin M. (2007) Dysphagia bedside screening for acute stroke patients. The gugging swallowing screen. *Stroke* **38**, 2948–2952.

Wang T.G., Chang Y.C., Chen S.Y. & Hsiao T.Y. (2005) Pulse oximetry does not reliably detect aspiration on videofluoroscopic swallowing study. *Archives of Physical Medicine and Rehabilitation* **86**(4), 730–734.

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

Westergren A. (2006) Detection of eating difficulties after stroke: a systematic review. *International Nursing Review* **53**, 143–149.

Wu C.H., Hsiao T.Y., Chen J.C., Chang Y.C. & Lee S.Y. (1997) Evaluation of swallowing safety with fiberoptic endoscope: comparison with videofluoroscopic technique. *Laryngoscope* **107**(5), 396–401.

Wu M.C., Chang Y.C., Wang T.G. & Lin L.C. (2004) Evaluating swallowing dysfunction using a 100-ml water swallowing test. *Dysphagia* **19**(1), 43–47.

Zenner P.M., Losinski D.S. & Mills R.H. (1995) Using cervical auscultation in the clinical dysphagia examination in long-term care. *Dysphagia* **10**(1), 27–31.

