



The Diagnostic Test Accuracy of Clinical Swallow Assessment for Oropharyngeal Aspiration: A Systematic Review

Marissa Romano BSpPath

Student number: a1619892

The Joanna Briggs Institute and School of Translational Health Science,
Faculty of Health Sciences, the University of Adelaide.

marissa.romano@adelaide.edu.au

Table of Contents

The Diagnostic Test Accuracy of Clinical Swallow Assessment for Oropharyngeal Aspiration: A Systematic Review.....	i
Table of Contents	ii
Table of Figures.....	v
Table of Tables	vi
Glossary of terms.....	vii
Acronyms.....	ix
Acknowledgments.....	x
Student Declaration	xi
Abstract	xii
1 Background	16
1.1 Systematic reviews	16
1.2 The systematic review protocol	17
1.3 Systematic review versus traditional literature review	18
1.4 Studies of diagnostic test accuracy	19
2 Introduction.....	22
2.1 Introduction to oropharyngeal aspiration.....	22
2.1.1 Definition of oropharyngeal aspiration.....	22
2.1.2 Neurology of swallowing and airway protection	22
2.1.3 Pathophysiology of chronic oropharyngeal aspiration	23
2.1.4 Causes and prevalence of oropharyngeal aspiration	24
2.1.5 Response to oropharyngeal aspiration in paediatrics compared with adults.....	27

2.1.6	Diagnostic testing for oropharyngeal aspiration	27
3	Systematic Review Protocol	31
3.1	The systematic review protocol	31
3.1.1	Background.....	31
3.1.2	Inclusion criteria	35
3.1.3	Focus of the review	37
3.1.4	Search strategy	38
3.1.5	Methods of the review	39
4	Results.....	41
4.1	Results.....	41
4.1.1	Results of the search	41
4.1.2	Methodological quality of the included papers	43
4.1.3	Results from individual QUADAS checklist items.....	46
4.1.4	Assessing methodological quality using the STARD checklist ..	52
4.1.5	Data from individual studies	58
4.1.6	Data analysis and synthesis.....	74
4.1.7	Meta-analysis.....	76
4.1.8	Sources of heterogeneity	80
4.1.9	Sub-group analysis	81
4.1.10	Summary of findings	83
5	Discussion	86
5.1	Discussion	86
5.1.1	Summary of main results	86
5.1.2	Findings of meta-analysis	86
5.1.3	Population and settings.....	87
5.1.4	Reference test methods.....	87

5.1.5	Index test methods.....	88
5.1.6	Effect of fatigue and aspiration.....	89
5.1.7	Heterogeneity of study results.....	89
5.1.8	The diagnostic test accuracy of elements of the CSA	94
5.1.9	Prevalence and positive and negative predictive values	96
6	Conclusions.....	99
6.1.1	Implications for practice	99
6.1.2	Concluding Summary.....	102
6.1.3	Implications for research	103
6.1.4	Limitations of the review.....	103
7	References	104
8	Appendices.....	114
8.1	Appendix I: Initial Search Terms.....	114
8.2	Appendix II: Electronic databases	115
8.3	Appendix III: PubMed Search Terms.....	116
8.4	Appendix IV: Critical appraisal tool – the QUADAS check list ⁸³	117
8.5	Appendix V: Data extraction tool – the STARD check list ⁸⁴	118
8.6	Appendix VI: Data extracted from individual studies and combined in meta-analysis	121
8.7	Appendix VII: Table of included studies.....	122
8.8	Appendix VIII: Table of excluded studies.....	132

Table of Figures

Figure 1 Flow chart outlining the study identification process.....	42
Figure 2 Methodological quality graph.....	44
Figure 3 Methodological quality item for each study included in meta-analysis.....	45
Figure 4 Forest plot of the included studies.	76
Figure 5 Summary receiver operator characteristic (sROC) plots of the sensitivity and specificity for all included studies.	78
Figure 6 Flowchart demonstrating the likelihood of a test result being accurate differs for a positive or negative test result.....	85

Table of Tables

Table 1 A typical 2x2 table used to classify patient test results and the presence or absence of disease or target condition.	20
Table 2 Differences between Bours review ⁸⁷ and this systematic review. ...	43
Table 3. Meta-analysis of diagnostic accuracy.	77
Table 4 A 2x2 classification table constructed using data extracted from the included studies.	80
Table 5 Summary of study findings.....	84
Table 6 Between study differences and similarities used to rate the presence or absence of aspiration during clinical swallow assessment and the cut-off values used to determine the score.	93
Table 7 Diagnostic accuracy of various features of CSA as presented in 5 of the 13 included studies..	95
Table 8 Comparison of PPV and NPV using varying prevalence of aspiration.	97
Table 9 A 2x2 classification used to determine PPV and NPV for prevalence of 60%	97
Table 10 A 2x2 classification used to determine PPV and NPV for prevalence of 10%.	98

Glossary of terms

Clinical Swallow Assessment: non-instrumental, non-radiologic assessment of swallow function.

Diagnostic test accuracy: the ability of a test to distinguish between patients with a target disease or condition from those without the disease or condition.

Dysphagia: difficulty in any of the four phases of swallowing.

False negative: index test result is negative, reference test result is positive.

False positive: index test result is positive, reference test result is negative.

Index test: the 'new' test or test in question, in a study of diagnostic test accuracy.

Oropharyngeal aspiration: the entry of food and/or fluids below the level of the vocal cords.

Quality Assessment of Diagnostic Accuracy Studies (QUADAS) checklist: 14 point checklist used to assess the methodological quality of studies of diagnostic test accuracy.

Reference test: the 'gold standard' test against which the index test is compared in a study of diagnostic test accuracy.

Sensitivity: the ability of a test to correctly identify those with the disease or target condition.

Specificity: the ability of a test to correctly identify those without the disease or target condition.

Standards for Reporting of Diagnostic Accuracy (STARD) checklist: 25 item checklist used to extract data of studies of diagnostic test accuracy.

Summary receiver operating characteristic plot: graphical representation used to describe the performance of a diagnostic test based on data from meta-analysis.

True negative: index test result and reference test result are negative.

True positive: index test result and reference test result are positive.

Video Fluoroscopic Swallow Study: radiographic assessment of swallow function.

Acronyms

CSA: clinical swallow assessment

CVA: cerebrovascular accident

DTA: diagnostic test accuracy

FEES: fiberoptic endoscopic evaluation of swallowing

FN: false negative

FP: false positive

QUADAS: Quality Assessment of Diagnostic Accuracy Studies

sROC plot: summary receiver operating characteristic plot

STARD: Standards for Reporting of Diagnostic Accuracy

TN: true negative

TP: true positive

VFSS: video fluoroscopic swallow study

Acknowledgments

I would like to thank the following people for the generosity of their time and expertise during the writing of this thesis: supervisors Dr Tim Schultz, Dr Andrew Tai, Dr Sarahlouise White, research librarian Maureen Bell, statistician Thomas Sullivan, second reviewer May Thwin, the Speech Pathology Department, Women's and Children's Hospital and Haruka Tohara, author of one of the included studies for providing the raw data clearly describing the number of participants in the study with aspiration present and aspiration absent.

Student Declaration

This work contains no material that has been accepted for the award of any other degree or diploma in any University of any other tertiary institution, and, to the best of my knowledge and belief, contains no material previously published or written by any other person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint award of this degree.

I give consent to this copy of my thesis, when deposited in the University Library, being made available for loan and photocopying, subject to the provisions of the Copyright Act 1968.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Catalogue, and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

Signed:

Dated:29/4/2014

Abstract

Background

Oropharyngeal aspiration, the recurrent entry of food and/or fluids below the level of the vocal cords, can result in a range of complications including: chronic lung diseases, aspiration pneumonia, malnutrition and/or dehydration. Video fluoroscopic swallow study is the Gold Standard assessment of oropharyngeal aspiration but is resource intense, exposes the patient to radiation and is not available in all hospitals and centres. The Clinical Swallow Assessment is a bedside swallow assessment widely used to screen and/or assess for oropharyngeal aspiration. The evidence base behind the diagnostic test accuracy of the Clinical Swallow Assessment has not previously been synthesised.

Objectives

To synthesise the best available evidence on the diagnostic test accuracy (sensitivity and specificity) of clinical swallow assessment compared with Video Fluoroscopic Swallow Study in diagnosing oropharyngeal aspiration in children and adults with dysphagia.

Inclusion criteria

Types of participants

Any patients referred for swallowing assessment, specifically assessed for oropharyngeal aspiration were included and there was no exclusion based on age or gender. Study results were excluded for head and neck cancer patients, patients with a tracheostomy *in situ* and patients with craniofacial anomalies.

Focus of the review

The focus of the review was to examine the diagnostic test accuracy of clinical swallow assessment, as compared with Video Fluoroscopic Swallow Study.

Types of studies

This systematic review considered any relevant cross sectional study that measured diagnostic test accuracy.

Types of outcomes

Outcomes of interest were the sensitivity and specificity of the clinical swallow, as compared with the video fluoroscopic study and the positive and negative predictive values. Where this data was not reported in the studies, these measures were calculated from the reported raw data.

Search strategy

Thirteen major databases were searched from their inception until April 31st 2012. There were no limits during the search stage as relevant studies were omitted if search filters such as 'English' and 'Human' were applied.

Methodological quality

Methodological quality was assessed using the QUADAS checklist. Data was collected using the STARD checklist. Sensitivity and specificity measures were combined in meta-analysis to generate a summary receiver operator characteristic plot.

Results

There were 1787 titles initially identified. Following duplicate removal and screening against inclusion criteria, 37 papers were retrieved for detailed examination and 24 papers were excluded as they did not meet the inclusion criteria. The most common reason for exclusion was that the paper was not a study of diagnostic test accuracy. There were 13 studies included in the systematic review and found to have high methodological quality. Data extracted from individual studies was statistically combined in meta-analysis to produce a forest plot and summary receiver operating characteristic (sROC) plot. Heterogeneity was evident in the forest plot, particularly for sensitivity as evidenced by the wider confidence intervals for sensitivity compared with specificity. The test sensitivity varied from 21% to 93%, the specificity from 46% to 93%. The summary mean sensitivity and specificity was calculated as 71% and 76% respectively. Positive predictive value was

calculated as 60% and negative predictive value was 81%. The scatter of points around the curve on the sROC plot also indicated heterogeneity. Sources of heterogeneity were identified and explored. The shape of the sROC curve strongly supported the finding of a threshold effect, which is expected for studies in which there is a strong interpretative component such as the clinical swallow assessment. This occurs as clinicians may vary in their criteria for what constitutes a positive or negative test result. The overall prevalence of aspiration in the included studies was calculated as 35%. Results are based predominantly on adult, acute post stroke patients.

Conclusion

This thesis provides good evidence for an overall estimate of the sensitivity and specificity of clinical swallow assessment compared with video fluoroscopic swallow study for the assessment of oropharyngeal aspiration. In this population, a clinician can be much more confident in a negative test result than a positive test result. A false positive test result may lead to unnecessary patient care and costs, including with-holding oral medications and prescription of modified diets and/or fluids. A false negative test result may lead to compromised lung health and/or pneumonia.

Implications for practice

Using calculations of the positive predictive values and negative predictive values, 60% of patients who test positive for aspiration are truly aspirating and 81% of patients who test negative for aspiration are truly not aspirating. Positive and negative test results are affected by the prevalence of the condition in the population. To summarise, the PPV increases and the NPV decreases as prevalence increases and the PPV decreases and NPV increases as the prevalence decreases. For example if the prevalence is much lower (e.g. 10%) the NPV rises to 96% and the PPV decreases to 24%. This thesis provides data for centres where VFSS is not available regarding the diagnostic test accuracy of clinical swallow assessment for oropharyngeal aspiration.

Implications for Research

Only one of the included studies provided data for infants and children. None of the included studies addressed infants, children or adults without a neurological aetiology. Further research is needed for infants and children with dysphagia as well as neurologically intact and normally developing infants, children and adults.

1 Background

1.1 Systematic reviews

A systematic review is defined as “a review of the evidence on a clearly formulated question that uses systematic and explicit methods to identify, select and critically appraise relevant primary research, and to extract and analyse data from the studies that are included in the review” p.3¹ and is considered the highest level of evidence.^{2, 3} Systematic reviews occupy this position because the review authors systematically search, identify and summarise the available evidence for a pre-defined, focussed clinical question.³ The review is a transparent process with a detailed account of the study identification process and the methodological quality of the included studies.³ Furthermore, systematic reviews are guided by a protocol that is finalised and published prior to commencing the review.

Systematic reviews are an essential component of health care research, providing a transparent, accurate and reliable summary of evidence.^{4, 5} Systematic reviews enable clinicians and policy-makers to access up to date current best practice and synthesised evidence necessary for policy-making decisions. Systematic reviews also provide a starting point for clinical practice guideline developers.⁵ Systematic reviews are becoming increasingly common, for example, recent data suggest 2,500 new systematic reports in English are indexed in MEDLINE annually.⁵

A systematic review of the literature involves the following seven steps:³

1. The development of a rigorous protocol (a systematic review protocol is included in the following section).
2. Stating the question/focus of the review.
3. Establishing criteria that will be used to select the literature.
4. Presenting a strategy that will be used to identify all relevant literature within a given time frame.
5. Stating how the quality of each study/paper will be assessed and any exclusion criteria.

-
6. Explaining how data will be extracted from the primary research or text.
 7. Describing how the extracted data will be synthesised.

It is the transparent and systematic process underpinning systematic reviews which allows the reader or user of the systematic review to evaluate the accuracy and appropriateness of the underlying methods used to present the relevant information. Systematic reviews are used within a range of health care areas such as identifying the clinical and/or cost effectiveness of an intervention or drug. ⁴ Systematic reviews are particularly insightful when there is a substantive clinical question and several primary studies present equivocal or disparate findings. ⁴ An example of this is cited in relation to the Cochrane library. ⁶ In this example, a single paper published in 1998 presented the results of 12 children presenting to a paediatric gastroenterology unit and suggested a possible link between the mumps, measles and rubella (MMR) vaccine and the development of Crohn's disease and Autism. ⁷ This resulted in a period of reduced uptake of the vaccine. ⁴ In 2004 a brief summary titled 'Retraction of an Interpretation' ⁸ was published by 10 of the 12 available authors of the original paper as the potential bias of this paper was recognised. ⁴ Following this a systematic review on this topic was published in the Cochrane Library ⁹ and in summary, assessed "no significant association between MMR immunisation and the following conditions: autism, asthma, leukaemia, hay fever, type 1 diabetes, gait disturbance, Crohn's disease, demyelinating diseases, or bacterial or viral infections" p.2 ⁹

1.2 The systematic review protocol

The protocol of a systematic review provides a pre-specified plan to ensure rigour and reduce potential bias. ^{3, 4} This pre determined plan, or protocol provides the reader with the background, rationale and intent against which the current literature will be appraised. The protocol also details the types of studies to be considered for inclusion in the review, as

well as what data will be extracted and how that data will be synthesised. The protocol is peer reviewed and published prior to commencing the search of all included databases and provides a transparent reporting of the available literature (published and unpublished) relevant to the pre-defined clinical question. By publishing the protocol prior to embarking on the systematic review, this provides a public record of the criteria against which the systematic review authors intend to address the stated research question.¹⁰ It also enables the review to be replicated, reproduced and updated as necessary.⁴

1.3 Systematic review versus traditional literature review

Systematic reviews are distinguished from other types of reviews in the literature such as literature reviews and research overviews, by their overarching aim of being more transparent and reproducible. Unlike systematic reviews, traditional literature reviews often use an informal, non-systematic approach to compile the published literature available for a given topic.¹ Furthermore traditional literature reviews do not necessarily state how the included studies are selected, appraised and contribute to the overall findings and recommendations of the review.⁴ Although traditional reviews can provide an insightful overview of a given topic, the methods used are not reproducible and the conclusions may not be valid,¹¹ leading to potential bias in the recommendations.⁴ Towards the late 1980s, the process used to generate literature reviews was examined closely and the inadequacies of their rigour was highlighted and explored.^{4, 11} Between June 1985 and June 1986, 50 medical reviews were appraised.¹¹ The following information was 'not specified' in more than 46 of the papers: data identification, data selection, validity assessment and quantitative synthesis. Other flaws of traditional literature reviews include: lack of explicit systematic methods to identify, assess and synthesise the information; broad purposes rather than a specific, focused research question; lack of standardised methodological criteria for assessing data

validity; lack of qualitative methods stated for data synthesis; poorly reported or absent information relating to future research implications. ¹¹

1.4 Studies of diagnostic test accuracy

Diagnostic test accuracy refers to the ability of a test to distinguish between patients with a target disease or condition from those without the disease or condition. ^{12, 13} In studies of diagnostic test accuracy the 'new' test or 'test in question' is referred to as the index test. The performance of the index test is compared to the reference or 'gold standard' test which is an established, accurate and agreed upon measure for the target condition or disease in question. ¹²⁻¹⁴ The index test may demonstrate additional benefits compared to the reference test and be considered for replacement of the reference test. For example, the index test may be less invasive, require less time to perform, require less specialist staff, be more cost effective, pose fewer risks to the patient, provide results in a shorter time period, be more readily available or be easier to interpret. ^{14, 15}

Two measures inherent to the accuracy of a test are sensitivity and specificity. The accuracy of the index test is reported in these measures relative to the reference test.

- Sensitivity refers to the ability of the test to correctly identify those with the disease or target condition. ¹⁶
- Specificity refers to the ability of the test to correctly identify those without the disease or target condition. ¹⁶

Patients are classified according to their test result. For example if the index test result is positive and the reference standard test result is positive, this is labelled a 'true positive' (TP) test result as both of the tests produced the same positive test result. However, if the index test result is positive but the reference standard test is negative, this is labelled a 'false positive' (FP) to highlight the discrepancy between the two test results. If the index test result is negative and the reference standard test result is negative, this is labelled a 'true negative' (TN) test result as both of the

tests produced the same negative test result. Conversely, if the index test result is negative but the reference standard test is positive this is labelled a 'false negative' (FN) to demonstrate the conflicting results of the index test compared with the reference test. ^{13, 16}

Table 1 presents a typical 2x2 table used to present test results. ¹³

Table 1 A typical 2x2 table used to classify patient test results and the presence or absence of disease or target condition.

Test outcome (Index Test Results)	Disease/condition status (Reference test results)		
	Disease/Condition positive	Disease/Condition negative	Total
Index test positive	True positives (TP) (a)	False positives (FP) (b)	Test positives (a+b)
Index test negative	False negatives (FN) (c)	True negatives (TN) (d)	Test negatives (c+d)
Total	Disease/condition positives (a+c)	Disease/condition negatives (b+d)	N (a + b + c + d)

Sensitivity and specificity can be calculated using the patient/subject data from Table 1. ^{13, 17}

Sensitivity:
$$\frac{a}{a + c}$$

Specificity:
$$\frac{d}{b + d}$$

Positive and negative predictive values can be calculated to assist with clinical decision making once the test result is known. ¹⁸ The positive predictive value (PPV) is relevant if the test result is positive, the negative predictive value (NPV) is relevant if the test result is negative. Positive predictive value refers to the proportion of patients with a positive test result who *truly* have the disease in question and the negative predictive value refers to the proportion of patients with a negative test result who

truly *do not* have the disease in question. ^{13, 16} For example a PPV of 90% indicates that 90% of people who received a positive test result following the index test, have the disease or target condition in question. PPV and NPV is influenced by disease prevalence. This is discussed in section 5.1.9 of this thesis.

Below are the formulae used to calculate PPV and NPV.

$$\text{PPV:} \quad \frac{\text{TP}}{\text{TP} + \text{FP}} \times 100$$

$$\text{NPV:} \quad \frac{\text{TN}}{\text{TN} + \text{FN}} \times 100$$

2 Introduction

2.1 Introduction to oropharyngeal aspiration

2.1.1 Definition of oropharyngeal aspiration

Oropharyngeal aspiration is the entry of saliva, fluid or food particles below the level of the true vocal cords.^{19, 20} The term 'oropharyngeal' aspiration is used to distinguish from other types of aspiration such as the aspiration of gastric content material.²¹ Oropharyngeal aspiration may occur in any infant, child or adult with dysphagia. Dysphagia is defined as difficulty in any of the four phases of swallowing.²² Swallowing in children and adults is divided into four phases: the oral preparatory, oral, pharyngeal and oesophageal.^{23, 24} Integrity of the pharyngeal phase of swallowing is of particular importance in preventing aspiration due to the anatomical placement of the pharynx in relation to the trachea. The pharynx shares the same space used during respiration which leads to the laryngeal inlet and upper airway.²⁵ Human beings are the only mammals with a shared space for breathing and swallowing which allows for humans to communicate using voice.²⁵

2.1.2 Neurology of swallowing and airway protection

As described above, swallowing may be divided into the oral, pharyngeal and oesophageal phase. The oral stage of swallowing is voluntary, that is the person has control over the biting, chewing and bolus manipulation. In contrast to this, the pharyngeal phase of swallowing is reflexive.²⁶ Once the swallow reflex is initiated, several events take place to ensure the bolus is transported into the oesophagus and the airway is protected. The primary protective mechanism to prevent aspiration during swallowing is the closure of the true vocal cords.²⁷ Following this, the false vocal cords and aryepiglottic folds adduct as the epiglottis deflects to assist with

directing the bolus towards the upper oesophageal sphincter. ²⁶ Following closure of the upper airway the larynx is elevated and moved anteriorly as pharyngeal peristalsis begins to transport the bolus through the pharynx and towards the upper oesophageal sphincter. ²⁸

Aspiration may occur before, during or after the swallow. In order to assist with treatment of aspiration and establish the aetiology, clinicians need to know at which point in the swallow the aspiration has occurred. ²⁵ However, this information is infrequently provided in the literature regarding diagnostic test accuracy for tests of aspiration as the focus is on the presence or absence of aspiration.

2.1.3 Pathophysiology of chronic oropharyngeal aspiration

The term 'chronic' aspiration is used to distinguish from a single acute event of aspiration. Chronic aspiration occurs when there is repeated entry of saliva, food or fluid below the level of the vocal cords. ²² Pulmonary aspiration is the term used to describe the entry of this material into the tracheobronchial tree. ²⁵ The tracheobronchial tree is the structure from the trachea, bronchi and bronchioles that forms the airways that supply air to the lungs.

Aspirated material may include food and liquids, oral secretions, mucous, breast milk (for breast feeding infants) and any oral medications. Several pulmonary events may occur in response to oropharyngeal aspiration depending on the amount and nature of the aspirated material, the frequency of the aspiration and the host's response to the aspirated material. ^{25, 29} The exact relationship between the volume and type of aspirated material and likelihood of developing pneumonia is not clear but there is some evidence to suggest a correlation between a larger volume of aspirate and the development of pneumonia in stroke patients. ³⁰ The patient's oral hygiene including condition of dentition can also contribute to the host's response to oropharyngeal aspiration. The oral cavity is a potential source of the organisms responsible for aspiration pneumonia,

with decayed teeth being significantly related to the incidence of aspiration pneumonia in patients with dysphagia.³¹

Aspiration may lead to an acute pulmonary inflammatory response to the bacteria and bacterial products.^{25, 29} Initially neutrophils then monocytes respond to these areas causing a foreign body reaction and granuloma formation. Long-term complications can include tracheal and bronchial granuloma formation, recurrent pneumonia, bronchiolitis and bronchiectasis.²⁵

2.1.4 Causes and prevalence of oropharyngeal aspiration

Oropharyngeal aspiration may occur if there is any anatomical or physiological disorder affecting one or all stages of the swallowing process.³² Dysphagia can occur in infants, children and adults. There is not one single aetiology of dysphagia, however, a large proportion of the published literature focuses on dysphagia following stroke or other neurological conditions.³³ Neurological conditions can affect the structures and precise neural control required for swallowing as well as the cough reflex which assists in airway protection in the event of aspiration.³⁴ Initially, a patient may not demonstrate signs of dysphagia. These signs may not be seen until part way through the feeding process due to the possible effects of fatigue on feeding and swallowing.³⁵ Including this potential 'fatigue' effect in the swallowing assessment is essential for accurate diagnosis and treatment.

Dysphagia is a common and serious sequela following stroke involving one or both cerebral hemispheres or the brain stem.³⁶ The number of adults presenting to clinicians with oropharyngeal dysphagia is increasing, in part due to the growing population of > 65 years and a longer life expectancy.³³ Despite the significant impact on patient's health and well-being, oropharyngeal dysphagia post stroke is underestimated and underdiagnosed as a cause of major nutritional and respiratory complications.³⁷ Prompt identification of oropharyngeal dysphagia post

stroke can be challenging, as it can occur with subtle, or absent associated neurological deficits.³⁸ Unilateral strokes lead to oropharyngeal dysphagia in 40% of patients, bilateral cerebral hemisphere lesions in 56%, brainstem lesions in 67% and combined lesions in 85%.³⁷ Recovery of oropharyngeal dysphagia post stroke is thought to be related to the neuroplasticity and compensation of the non-affected hemisphere.³³

Other neurological conditions that may result in dysphagia in adults include: brain injury, spinal cord injury, Parkinson's disease, multiple sclerosis, cerebral palsy and Alzheimer's disease.³⁹ Non neurological conditions include head and neck cancer.⁴⁰ Parkinson's disease is a bradykinetic disorder with an incidence of approximately 13 in 100,000.³³ Prevalence rates of dysphagia in this population are 82% with aspiration pneumonia the most common cause of death.³³ Multiple sclerosis is a demyelinating immune-mediated condition. The incidence of oropharyngeal dysphagia in people with multiple sclerosis is documented as varying from 24% to 65%, depending on the severity of the condition.³³ Similarly to people following stroke, oropharyngeal dysphagia is most severe in patients with brain stem pathology.³³ The presence of dysphagia in neurologically intact/otherwise normal adults is not well documented in the literature, with the exception of the reasonably well described age-related dysphagia.^{39, 41}

Dysphagia in infants and children can either exist alone or in addition to other underlying medical conditions. Dysphagia is well documented in infants and children in the following patient groups: prematurity, upper aero-digestive tract anomalies, central nervous system impairments and neurodevelopmental delays.^{42, 43} More recently, dysphagia has been identified in infants and children without known risk factors associated with swallowing difficulties.⁴²

There is a reasonable amount of variability regarding data for the incidence and prevalence of dysphagia amongst adults and children possibly due to the differences in study design and populations included in the studies. One systematic review paper reported the incidence for

dysphagia in studies enrolling acute adult stroke patients ranged from 64% to 78% and aspiration ranging from 22% to 50%.⁴⁴ The incidence figures for feeding-related difficulties in children vary considerably, which is likely due to the various diagnostic labels used in the literature.⁴⁵ At present there is no universally accepted classification for paediatric feeding disorders^{46, 47} which impacts on data collection and incidence/prevalence estimates.

The incidence of feeding disorders is estimated to be between 25 – 45% in typically developing children and up to 80% in children with developmental disabilities. Incidence may vary between studies due to the method used to assess the presence or absence of dysphagia, the age of the population included and patient demographics. For example, one study in the USA⁴⁸ reported an incidence of 35% of feeding difficulties for a population of ‘normally developing’ infants aged 0 – 4 months. A parent questionnaire was used to identify these feeding difficulties using four specific categories: excessive crying, spitting, colic and feeding difficulties. A second study in New Zealand⁴⁹ reported an incidence of 25%. This study was based on parent report of typically developing children, two years of age.

Children with developmental disabilities have a high reported prevalence of dysphagia.⁴⁵ In particular, children with central nervous system conditions such as cerebral palsy.^{47, 50} A recent study demonstrates that as the severity of gross motor function in cerebral palsy patients increased, so did the presence of swallowing difficulties.⁵¹ Previous studies have similarly shown the prevalence of swallowing difficulties in children with hemiplegia or diplegia is less than in children with spastic quadriplegia or extrapyramidal cerebral palsy.⁴⁵

The incidence of paediatric dysphagia as opposed to ‘feeding disorders’ is unknown,⁴⁵ however, the literature suggests the incidence is increasing.^{45, 51, 52} Factors contributing to this rise in incidence of paediatric dysphagia include improved survival rates for: pre-term infants, low birth weight and very low birth weight infants and the survival of infants born

with complex medical conditions.^{45, 51} Furthermore, the life expectancy of children with cerebral palsy is also increasing.⁵³

2.1.5 Response to oropharyngeal aspiration in paediatrics compared with adults

Responses to aspiration are different in children compared with adults. For example causes of dysphagia and aspiration in children include: prematurity, gastro-oesophageal reflux, neurological disorder and respiratory difficulties^{34, 39} and symptoms suggestive of aspiration include: upper airway noises, apnoea and cyanosis with feeds.⁵⁴ Oropharyngeal dysphagia in children without known neurological conditions and without apparent risk factors is not well documented in the literature and few clinical guidelines are available for assessment and treatment in this group.⁵⁵ These children may present as neurologically intact and normally developing children with ambiguous or unexplained respiratory symptoms.⁵⁵

In adults, symptoms suggestive of aspiration include: coughing, extra chewing time, holding food or fluid in the mouth and in some cases loss of food or fluid from the mouth.³⁹

2.1.6 Diagnostic testing for oropharyngeal aspiration

The 'Gold Standard' for assessment and diagnosis of oropharyngeal aspiration is the Video Fluoroscopic Swallow Study (VFSS).^{22, 56} The terms VFSS and Modified Barium Swallow (MBS) refer to the same technique.^{34, 57} VFSS is the radiographic assessment of swallow function with a focus on the oral, pharyngeal and upper oesophageal phase of swallowing. It is frequently used as the reference standard in tests of diagnostic test accuracy.^{20, 42, 58, 59} However, VFSS exposes the patient to radiation, is relatively expensive, requires specialist equipment and staff and is not available to all clinicians.⁵⁹ VFSS is not available in all hospitals

and centres or the equipment may be available without appropriately trained staff to perform the test.

In contrast, the Clinical Swallow Assessment or 'Bedside Swallow Assessment' is a non invasive assessment of swallowing and oral feeding skills and is widely available to all clinicians who have undergone the necessary training to perform this assessment.⁶⁰ Training requirements for staff performing this assessment vary between hospitals and centres. The terms Clinical Swallow Assessment and Bedside Swallow Evaluation also refer to the same diagnostic tool.⁶¹ As VFSS is not available in all settings, clinical swallow assessment may be the only available swallowing test for patients with dysphagia.

Video Fluoroscopic Swallow Study

The use of fluoroscopy for assessment of swallowing was first mentioned in 1898.⁶² At that time the purpose was to assess the oesophageal phase of swallowing. In 1927, Mosher, a laryngologist, used fluoroscopy to further assess features of the pharyngeal swallow. Over time, new equipment allowed for a more detailed observation and record of swallow function⁶³ and studies included a wide range of subjects - infants, children and adults with and without medical co morbidities.

The use of VFSS has evolved since this time and the first published protocol for VFSS was by Logeman in 1983.⁶⁴ VFSS is now used as a diagnostic tool to assess airway protection in patients with clinical symptoms suggestive of aspiration. VFSS allows the clinician to objectively evaluate all phases of swallowing and diagnose aspiration and silent aspiration.⁴² The use of VFSS also enables diagnosis of post swallow residue within the pharynx which has been linked to pneumonia.²⁰ VFSS allows the clinician to trial consistencies and altered patient positioning to establish an eating and drinking program without aspiration.²¹ In addition the VFSS films may be recorded and replayed to families to assist with relaying the results and recommendations of the study.⁶⁵ The study aims to identify fluids that do and do not result in aspiration. The results are presented as a binary outcome whereby the clinician records

whether the infant, child or adult is aspirating.

One of the main disadvantages of the VFSS is the patchiness of its availability due to the specialist equipment, specialist staff and training required.⁶⁶ VFSS is also relatively resource intense due to this specialist equipment and staffing required to operate the equipment and interpret the study findings.⁶⁶ The fluids and solids used during the study are radio-opaque and prepared with a contrast agent. These fluids are intended to represent the infant's formula or patient's fluid, however, studies have shown the fluids used during VFSS are not necessarily representative of the viscosity of fluid being tested.^{58, 67} For example, a 'thin' fluid recipe using barium and water may have a thicker viscosity than normal water usually consumed by the patient.

VFSS does expose the patient to radiation. In studies measuring the radiation dose to patients, the value is expressed in milliSieverts (mSv) – a unit used to measure radiation dose to the body. An 'effective dose' level is provided which is a measure of the overall risk to the patient from radiation exposure. The radiation dose is influenced by several factors including: patient height and weight, patient position, equipment used and the overall screening time.^{68, 69} Experienced clinicians use less fluoroscopy time than novice clinicians which has a direct impact on the patient's overall radiation exposure.⁷⁰ A recent study addressing radiation exposure to children during VFSS reported effective doses in the range 0.01-0.25 mSv with a mean effective dose of 0.08 mSv.⁶⁹ These results are well below the dose limit of 1 mSv per year for public exposure suggested by Australian Radiation Protection and Nuclear Safety Agency.⁶⁹ A similar study reviewing adults undergoing a VFSS reported a median of 0.85 mSv. This can be compared to a single postero-anterior chest x-ray (0.02 mSv), a postero-anterior and lateral chest x-ray (0.1 mSv) or a computed tomographic (CT) chest which is a much higher dose (7.00 mSv).⁶⁸

Clinical Swallow Assessment (CSA)

Clinical Swallow Assessment (CSA) is performed by a speech pathologist in any setting including the home or outpatient clinic. For the purpose of this study CSA is defined as the assessment of swallowing by a speech pathologist or occupational therapist using fluids and solids of varying viscosity. It usually involves a thorough case history, cranial nerve assessment and review of the patient eating and drinking their usual food and fluid.^{59, 71} Speech pathologists use a range of signs and symptoms during clinical swallow assessment as indicators of dysphagia including cough, wheeze, and respiration and voice changes.⁵⁴ The assessment guides recommendations for 'safe swallowing' and the speech pathologist will often recommend a modified diet, fluids or feeding strategies. Clinical swallow assessment is inexpensive and does not require additional specialist staff (such as radiographer and radiologist) or specialist equipment. It can be repeated frequently for patients with rapidly changing dysphagia.⁷² However, CSA does rely on the skills and experience of the clinician performing the test.

Coughing or signs of choking are commonly used as clinical markers to diagnose the presence of aspiration. However, some studies have shown clinical swallow assessment is a poor diagnostic tool for assessment of aspiration compared with VFSS²² particularly for patients experiencing silent aspiration, which is common^{22, 42, 71}, particularly in children with neurological impairment.⁷³ Therefore, silent aspiration may reduce the reliability of clinical swallow assessment as a diagnostic tool for aspiration.⁷³ In contrast, other studies have shown CSA is a reasonable screening tool for children and adults at risk of oropharyngeal aspiration⁷⁴, particularly aspiration of fluids.⁵⁹ Nevertheless, CSA is an essential diagnostic tool for centres where VFSS facilities are not available.

3 Systematic Review Protocol

3.1 The systematic review protocol

The following chapter is the published protocol for this systematic review.

⁷⁵ The format is based on recommendations by the Joanna Briggs Institute and consists of standardised sections and includes some material from the previous chapter (Chapter 2) as background to the review. The protocol is available from the Joanna Briggs Database of Systematic Reviews and Implementation Reports <http://www.joannabriggslibrary.org/jbilibrary>

3.1.1 Background

Swallowing is a complex and dynamic process. It involves the precise co-ordination of over 31 paired muscles, six cranial nerves and the central nervous system including the brain stem and cerebral cortex. ^{23, 25, 42, 76}

Swallowing in children and adults is divided into four phases: the oral preparatory, oral, pharyngeal and oesophageal. ^{23, 24} Disruption to the normal sequence of swallowing during any or all of these phases is termed dysphagia and is defined by American Speech-Language-Hearing Association as “A swallowing disorder. The signs and symptoms of dysphagia vary and may involve the mouth, pharynx, larynx, and/or esophagus”. ⁷⁷ Dysphagia can affect infants, children and adults.

The pharyngeal phase of swallowing is of particular interest as it involves structures of the larynx and pharynx which are also used during respiration (breathing) and phonation (talking). ²⁵ During the normal process of swallowing, the trachea is protected as the bolus passes through the pharynx and enters the oesophagus. ⁵⁶ However, patients with dysphagia may experience penetration or aspiration. Penetration is the entry of fluid or food particles into the laryngeal vestibule which does not pass below the level of the true vocal cords. ¹⁹ Aspiration is the entry of

fluid or food particles below the level of the true vocal cords.^{19, 20} Silent aspiration is the term used when aspiration occurs without a cough response.

Prompt and accurate diagnosis of aspiration is integral to patient management. It enables immediate intervention to provide the patient with sufficient nutrition and hydration without the risk of aspiration lung disease. If aspiration is not accurately diagnosed, appropriate interventions cannot be implemented. Aspiration left untreated can have a range of poor health outcomes including: inadequate nutrition and hydration, compromised lung development and pulmonary integrity,²² chronic lung disease⁴² and recurrent respiratory symptoms.²¹

In infants and young children, aspiration episodes may be associated with transient cyanosis, recurrent chest infections or pneumonia.⁷⁸ Prompt diagnosis and treatment may also alleviate further aspiration-related damage to lung growth or function. The diagnosis of aspiration in young infants may be the initial presenting feature of an airway anomaly such as laryngeal cleft⁷⁸ or a neurological disorder. The diagnosis of aspiration may guide direction for further necessary investigations or interventions.

The swallowing pattern of infants and children is different to adults and the anatomy and physiology of swallowing continues to change from infancy through to adulthood.^{34, 79} The anatomical differences in an infant provide optimal conditions for breast feeding by allowing the infant to use a combination of jaw, cheek and tongue movement to create the sucking rhythm required for breast feeding. The position of the larynx is 'higher' than in an older child or adult and provides optimal airway protection during swallowing. As the infant grows the larynx descends and begins to resemble the anatomy of a young child by around 6 months of age.³⁴ By three years of age children begin to adopt a swallowing pattern similar to an older child.²⁵

Causes of dysphagia and responses to aspiration are also different in children compared with adults. For example causes of dysphagia in children include: prematurity, reflux, neurological disorder and respiratory

difficulties ^{34, 39} and symptoms suggestive of aspiration include: upper airway noises, apnoea and cyanosis with feeds. ⁵⁴ Causes of dysphagia in adults include: stroke, Parkinson's disease and Alzheimer's disease ³⁹ and symptoms suggestive of aspiration include: coughing, extra chewing time and loss of food or fluid from the mouth. ³⁹

The presence of neurological conditions may impact on the incidence, symptomatology and diagnosis of oropharyngeal aspiration. Oropharyngeal dysphagia is well documented in people with neurological conditions, (for example cerebral palsy and post-stroke) and such conditions predisposes them to oropharyngeal aspiration. ^{34, 40, 42} These conditions can affect the structures and precise neural control required for swallowing. ³⁴ The cough reflex, in response to aspiration may also be affected. ³⁴

Oropharyngeal dysphagia in children without known neurological conditions and without apparent risk factors is not well documented in the literature and few clinical guidelines are available for assessment and treatment in this group. ⁵⁵ These children may present as normally developing children, ⁵⁵ with ambiguous or unexplained respiratory symptoms. ⁴²

Swallowing Assessments

The 'Gold Standard' for assessment and diagnosis of oropharyngeal aspiration is a Video Fluoroscopic Swallow Study (VFSS) ^{22, 56} used frequently as the reference standard in tests of diagnostic test accuracy. ^{20, 42, 58, 59} VFSS exposes the patient to radiation, is relatively expensive, requires specialist equipment and staff and is not available to all clinicians. ⁵⁹ In contrast, the Clinical Swallow Assessment or 'Bedside Swallow Assessment' is a non invasive assessment of swallowing and oral feeding skills and is widely available to all clinicians.

Video Fluoroscopic Swallow Study

VFSS is the radiographic assessment of swallow function with a focus on the oral, pharyngeal and upper oesophageal phase of swallowing. The use of fluoroscopy for assessment of swallowing was first mentioned in

1898.⁶² At that time the purpose was to assess the oesophageal phase of swallowing. In 1927, Mosher, a laryngologist, used fluoroscopy to further assess features of the pharyngeal swallow. Over time, new equipment allowed for a more detailed observation and record of swallow function⁶³ and studies included a wide range of subjects - infants, children and adults with and without medical co morbidities.

The use of VFSS has evolved since this time and the first published protocol for VFSS was by Logeman in 1983.⁶⁴ VFSS is now used as a diagnostic tool to assess airway protection in patients with clinical symptoms suggestive of aspiration. VFSS allows the clinician to objectively evaluate all phases of swallowing and diagnose aspiration and silent aspiration.⁴² The use of VFSS also enables diagnosis of post swallow residue within the pharynx which has been linked to pneumonia.²⁰ VFSS allows the clinician to trial consistencies and altered patient positioning to establish an eating and drinking program without aspiration.²¹ In addition the VFSS films are recorded and able to be replayed to families which has been shown to increase compliance with clinician recommendations.

The fluids and solids used during the study are radio-opaque, prepared with a contrast agent. These fluids are intended to represent the infant's formula in order to establish fluids that do and do not result in aspiration. The results are presented as a binary outcome whereby the clinician records that the infant, child or adult is aspirating or is not aspirating.

Clinical Swallow Assessment (CSA)

Clinical Swallow Assessment (CSA) is performed by a speech pathologist and the patient in a natural setting such as home or outpatient clinic. It usually involves a thorough case history, cranial nerve assessment and review of the patient eating and drinking their usual food and fluid.^{59, 71} Speech pathologists use a range of signs and symptoms during clinical swallow assessment as indicators of dysphagia including cough, wheeze and voice changes.⁵⁴ The assessment guides recommendations for 'safe swallowing' and the speech pathologist will often recommend a modified diet, fluids or feeding strategies. Clinical swallow assessment is

inexpensive and does not require additional specialist staff (such as radiographer and radiologist) or specialist equipment. It can be repeated frequently for patients with rapidly changing dysphagia. ⁷²

For the purpose of this study, CSA is defined as the assessment of swallowing by a speech pathologist or occupational therapist using fluids and solids of varying viscosity .The results are recorded using a pre-determined checklist which includes variables frequently considered for assessment of swallow function such as: respiration changes, voice changes and presence or absence of coughing. ⁵⁹ The results are presented as a binary outcome.

Some studies have shown clinical swallow assessment is a poor diagnostic tool for assessment of aspiration ²² particularly for patients experiencing silent aspiration. Coughing or signs of choking is commonly used as a clinical marker to diagnose the presence of aspiration, however silent aspiration is common ^{22, 42, 80} particularly in children with neurological impairment. ⁷³

This may reduce the reliability of clinical swallow assessment as a diagnostic tool for aspiration. ⁷³ In contrast, other studies have shown CSA is a reasonable screening tool for children and adults at risk of oropharyngeal aspiration, ⁷⁴ particularly aspiration of fluids. ⁵⁹ Clinical swallow assessment is an essential diagnostic tool for centres where VFSS facilities are not available.

A preliminary search of JBI Library of Systematic Reviews, JBI COonNECT+, The Cochrane Library, PubMed and CINAHL has been conducted and revealed that no other systematic review either exists on this topic or is under way.

3.1.2 Inclusion criteria

The following are the inclusion criteria against which identified titles will be considered for inclusion in the review.

Types of participants

Studies will be considered for inclusion in this review if they include: infants, children or adults with dysphagia.

The following definitions will be used within this review:

- infant will be defined as < 36 months;
- children will be defined as 36 months – 18 years;
- adults will be defined as 18 years and over;
- dysphagia will be defined as 'a swallowing disorder'.⁷⁷

The above definition uses 36 months as the cut off between infants and children as this is the stage when the infant transitions from sucking patterns used for bottle feeding and early solids and begins to chew solids, self feed and drink from a range of cups consistent with an older child.²⁵

There will be no exclusion of studies based on age or gender of participants.

Studies will be excluded that: investigate head and neck cancer (HNC) patients, patients with a tracheostomy in situ and patients with craniofacial anomalies. However, if less than 10% of the study participants included one of these diagnoses the study will be considered for inclusion. The primary treatment for HNC patients includes surgery and radiotherapy.⁵⁷ Radiotherapy can have a specific impact on swallowing,⁸¹ not necessarily seen in other populations and surgical resections can result in predictable swallowing difficulties.⁸¹ Patients with a tracheostomy in situ and patients with craniofacial anomalies also have known anatomical changes that have a specific impact on swallow function.

Additionally, the clinician undertaking the clinical swallow assessment has access to detailed information on anatomical changes following surgery. This is different to other populations whereby information regarding specific neurological and anatomical information is not necessarily available.

3.1.3 Focus of the review

The focus of this review will be on the diagnostic test accuracy (sensitivity and specificity) of Clinical Swallow Assessment (CSA) compared with Video Fluoroscopic Swallow Study (VFSS) in diagnosing oropharyngeal aspiration in children and adults with dysphagia. In this review, VFSS will be the 'Gold Standard' or, reference test and CSA will be the index test.

The terms VFSS and Modified Barium Swallow (MBS) refer to the same type of technique.^{34, 57} Although this review will use the term VFSS defined as the radiographic assessment of the oral and pharyngeal swallow, MBS studies will also be included.

The terms Clinical Swallow Assessment and Bedside Swallow Evaluation also refer to the same diagnostic tool.⁶¹ Clinical Swallow Assessment (CSA) will be used within this review, defined as the non-instrumental, non-radiologic assessment of swallow function by a Speech Pathologist, however, studies of Bedside Swallow Evaluation will also be included.

Studies that do not use CSA as the index test or do not use VFSS as the reference test will be excluded. In addition, studies of effectiveness, experience or comparison of treatment interventions will also be excluded.

The diagnostic test accuracy of CSA and VFSS will be compared using: sensitivity and specificity. Where possible, positive and negative predictive values will also be analysed and reported.

Sensitivity of a diagnostic test is defined as 'the ability of the test to identify correctly those who have the disease'.¹⁶

Specificity of a diagnostic test is defined as 'the ability of the test to identify correctly those who do not have the disease'.¹⁶

Positive predictive value is defined as the proportion of patients who test positive and actually have the disease in question.¹⁶

Negative predictive value is defined as the proportion of patients who test negative and actually do not have the disease in question.¹⁶

Types of studies

This review will include studies of diagnostic test accuracy. It is anticipated studies will predominantly be cross-sectional studies.

3.1.4 Search strategy

The search strategy aims to find all relevant available literature, published and unpublished. Initial search terms and databases were considered by researching this area ^{13, 82} and discussion with a research librarian. Databases will be searched from their inception, to April 31st 2012. A three-step search will be used. Initially, a limited search of PubMed and CINAHL will be undertaken in order to identify appropriate keywords. Analysis of the text words and MeSH terms identified by the search to describe relevant articles will then be used to identify additional search terms which will then be used to search across all included databases. The initial search terms are listed in Appendix I and the databases to be searched are listed in Appendix II. Thirdly, the reference list of identified papers will be searched for additional studies. Hand searching of relevant journals will also occur for the following journals – Dysphagia and International Journal of Speech-Language Pathology.

The search will not be limited by year but will be limited to those published in the English language.

The following list of PubMed MeSH terms are also used to describe VFSS and CSA and will be included in the search strategy

- deglutition: The act of taking solids and liquids into the gastrointestinal tract through the mouth and throat.
- pneumonia, aspiration: A type of lung inflammation resulting from the aspiration of food, liquid, or gastric contents into the upper respiratory tract.
- photofluorography: The photography of images produced on a fluorescent screen by X-rays.

-
- videofluorography: Motion picture study of successive images appearing on a fluoroscopic screen.

3.1.5 Methods of the review

Assessment of methodological quality

Selected studies will be assessed by two independent reviewers for methodological validity prior to inclusion in the review using the QUADAS checklist ⁸³ (Appendix IV). Any disagreements that arise between the reviewers will be resolved through discussion, or with a third reviewer.

Data collection

Data will be extracted from included studies using the STARD checklist ⁸⁴ consisting of 25 items (Appendix V). The data extracted will include specific details regarding: populations, data collection and methods used for calculating or comparing measures of diagnostic accuracy. In situations where relevant study features are not provided within the study paper, the reviewer will contact the author to attempt to source additional information.

Data synthesis

The sensitivities and specificities from individual studies will be combined to generate a summary estimate of the accuracy of clinical swallow assessment compared with VFSS. Meta-analysis includes graphing the results of individual studies. The sensitivity and specificity are plotted as points on a graph. These plotted points are presented as an sROC (summary receiver operating characteristic) space, demonstrating the covariation between sensitivity and specificity. ¹² Revman 5 (Cochrane Collaboration) ¹³ and Open Office.org Calc computer software will be used for data analysis and synthesis.

Where possible, subgroup analyses will include:

- infants, children and adults

-
- subjects with a neurological condition compared to subjects without a neurological condition

Consensus does not exist for age ranges that define infancy, childhood and adolescence and definitions vary depending on the type of research being conducted.⁸⁵

Study results will be examined graphically, not statistically for heterogeneity and threshold effects.⁸⁶ Statistical analysis of heterogeneity is not possible for this data. Where statistical pooling is not possible the findings will be presented in narrative form including tables and figures to aid in data presentation where appropriate.

4 Results

4.1 Results

4.1.1 Results of the search

Using the 13 databases listed in Appendix II and the search strategy detailed in Appendix III to search, the 1787 potentially relevant titles were identified. 1033 of these were removed as they were obvious duplicates, identified using the EndNote computer software duplicate identification process as well as manual searching and removal of duplicates. Although the search included the terms 'sensitivity' and 'specificity' several descriptive studies were identified in the search. These descriptive studies presented the clinical relevance of clinical swallow assessment and/or video fluoroscopic swallow study, however they were not studies of diagnostic test accuracy and therefore not relevant to the review. Several studies were published in languages other than English and a very small number involved animals. The remaining 754 potentially relevant papers were screened using words in the title, keywords and abstract and compared against the inclusion criteria. On the basis of this screening procedure, 717 papers were excluded and 37 papers were retrieved for detailed examination and consideration for inclusion in the review. 24 studies were excluded following full text retrieval predominantly as the papers were not studies of diagnostic test accuracy. The remaining 13 studies were critically appraised by two independent reviewers and all 13 studies were included (see Appendix IV for the critical appraisal tool). No additional studies were identified for inclusion from the reference lists of studies retrieved for full text review. Figure 1 details the study identification process. See Appendix VII for table of included studies and Appendix VIII for table of excluded studies.

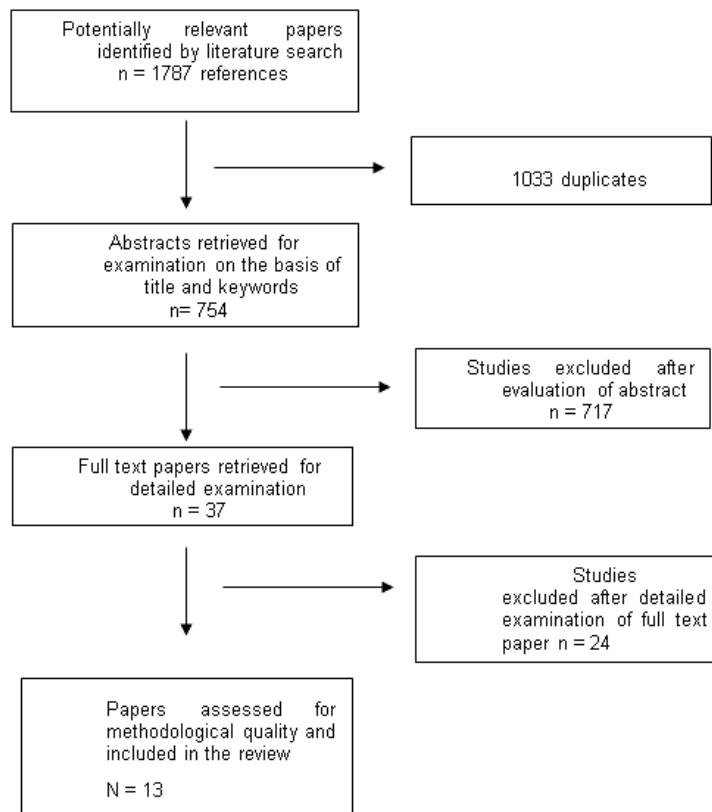


Figure 1 Flow chart outlining the study identification process.

One paper was identified in the search stage of this systematic review, titled “Bedside screening tests vs. video fluoroscopy or fibre-optic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review”.⁸⁷ This review was not identified during the preliminary search for existing systematic reviews on this particular topic, and was identified once the current review was underway. At first glance there were some concerns that the Bours review had similar aims to the present study. Following careful review it became apparent that there were important differences. In summary, the main differences between the Bours paper and this systematic review included the study aim, the population and the reference stand used. The aim of the Bours paper was to determine the effectiveness of bedside swallow assessment to detect dysphagia not aspiration, the population included only adult patients with a neurological condition and the reference standard included either VFSS or FEES (fibre-optic endoscopic evaluation of swallowing).

The differences between the Bours review ⁸⁷ and this systematic review are described in the table below.

Table 2 Differences between Bours review ⁸⁷ and this systematic review.

Bours review	This systematic review
Aim was to determine effectiveness of bedside methods to detect dysphagia, not aspiration.	Aim was to evaluate diagnostic test accuracy of clinical swallow assessment to detect aspiration.
End point for the reference test was penetration or aspiration.	End point for the reference test was aspiration.
Used two possible reference standard tests either FEES (fiberoptic endoscopic evaluation of swallowing) or VFSS.	Used VFSS as reference standard test.
Reported sensitivity and specificity for various features of clinical swallow assessment but clinician does not report whether aspiration is present or absent.	Following clinical swallow assessment clinician reports whether aspiration is present or absent.
Included use of oximeter in clinical swallow assessment.	Did not include any additional tools in clinical swallow assessment such as oximeter.
Included patients only with a neurological condition.	Included patients with and without neurological conditions.
Included adult population only.	Included infants, children and adults.

4.1.2 Methodological quality of the included papers

Critical appraisal using the QUADAS checklist

The QUADAS checklist ⁸³ was used to assess the quality of the conduct and reporting of included studies. The checklist items relate to known areas of bias and are recognised indicators of methodological quality for diagnostic test accuracy studies. ⁸³ All studies appraised for their methodological quality were included in this systematic review. There was agreement between the two reviewers, therefore a third reviewer was not required. The following section details how included studies performed against the 14 individual checklist items. Figure 2 and Figure 3 summarise the methodological quality of the included studies against the central themes of the QUADAS checklist. ⁸³ Figure 2 shows the overall quality of the 13 included studies as described by the ten central themes of the QUADAS checklist and Figure 3 shows the quality assessment results for individual studies as measured by the 14 individual items included in the QUADAS checklist. ⁸³

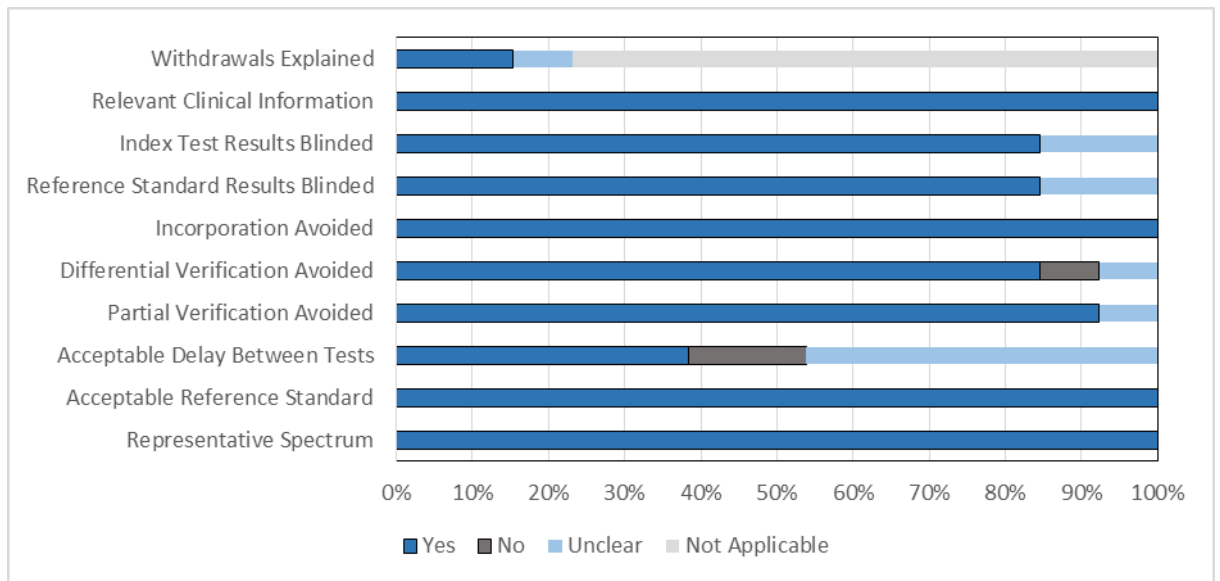


Figure 2 Methodological quality graph: Methodological quality items presented as percentages across all studies included in meta-analysis (n=13)

The overall results of quality appraisal are presented in Figure 2 and provide a summary of the validity of the available evidence because criteria that are unclear or not met introduce the risk for bias. As demonstrated in Figure 2 the overall strengths of the included studies were: providing all relevant clinical information such as details of the index test and reference standard, clinicians were blinded to test results, an acceptable reference standard was used as well as a representative patient spectrum. The three potential biases relating to the reference standard were avoided: incorporation bias – when the index test is incorporated into the reference standard; differential verification – when a set of patients is verified with a second or third reference standard; partial verification – when a non-random set of patients does not undergo the reference standard.

The main factor identified as a source of potential bias for the included studies was the time delay between tests which was marked as ‘no’ or ‘unclear’ in more than half of the included studies. In addition, a large number of studies scored ‘n/a’ regarding if withdrawals were explained.

	Withdrawals explained?	Sufficient detail to replicate index test?	Sufficient detail to replicate reference standard?	Clearly described selection criteria?	Reference standard independent of index test?	Index test results blinded?	Reference standard results blinded?	Whole or random sample receive reference standard?	Same reference standard?	Same clinical data available when test used in practice?	Uninterpretable or indeterminate results reported?	Acceptable delay between tests?	Acceptable reference standard?	Representative spectrum ?
Barbiera et al. 2006	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Baylow et al. 2009	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
DeMatteo et al. 2005	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
DePippo et al. 1992	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Hammond et al. 2009	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Linden et al. 1993	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?	<input checked="" type="checkbox"/>	?	?	<input checked="" type="checkbox"/>	?	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Mann et al. 2000	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
McCullough et al. 2001	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
McCullough et al. 2005	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Splaingard et al. 1988	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Smithard et al. 1998	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Tohara et al. 2003	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>
Zhou et al. 2011	n/a	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>	n/a	?	<input checked="" type="checkbox"/>	<input checked="" type="checkbox"/>

= yes; = no; ?= unclear; n/a= not applicable

Figure 3 Methodological quality item for each study included in meta-analysis.

Each of the included studies were appraised using the QUADAS checklist. Figure 3 illustrates the results for each of the 14 quality items in an abbreviated form. The QUADAS checklist can be found in its entirety in Appendix IV. This figure illustrates the strengths across all included studies were the inclusion of a representative patient spectrum, clearly described selection criteria, use of acceptable reference standard and use of reference standard independent of the index test. An overall weakness of the included studies was an unacceptable or unclear time delay between tests. Nine studies scored 11 or more using the QUADAS checklist (14 items). Two scored 10, one scored 9 and one scored 5. Although QUADAS provides a score out of 14, questions 13 and 14 on the checklist were often marked as 'n/a' for these included papers.

4.1.3 Results from individual QUADAS checklist items

The following section provides a narrative description of each of the 14 QUADAS Checklist items for the included studies.

1. Was the spectrum of patients representative of the patients who will receive the test in practice?

All of the 13 studies included in meta-analysis included patients presenting with dysphagia and referred for swallowing assessment which is representative of the patients who will receive the test in practice. Several studies specified the time post onset of diagnosis resulting in dysphagia.

2. Were selection criteria clearly described?

Two of the included studies did not clearly describe the selection criteria for inclusion into the study.^{88, 89} Four studies stated the patients were referred for swallowing assessment/evaluation,^{59, 90-92} two studies included patients with a diagnosis of stroke within 6 weeks^{61, 93}, one study included patients presenting to hospital within 24 hours of stroke⁹⁴, one study was an inception cohort of all patients presenting with first ever

stroke⁹⁵ one study described the patients only as presenting with a stroke⁹⁶ one study included resident patients on a stroke rehabilitation unit with one or more pre specified features indicating possible dysphagia (difficulty swallowing)⁹⁷ and one study described the included participants as experiencing symptoms or signs of dysphagia.⁹⁸

3. Is the reference standard likely to correctly classify the target condition?

VFSS is considered the best available, 'gold standard' test for assessment of aspiration.^{22, 56} Only studies which used this as the reference test were considered for inclusion in this systematic review.

4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?

Consideration of a reasonable time period between the two tests partially depends on the time post onset of dysphagia. For example, signs of aspiration are likely to rapidly change within the first 48 hours of a recent diagnosis of dysphagia compared with a person experiencing dysphagia for a longer period of time. The following guidelines were used to score this item: scored 'yes' if tests occurred within 48 hours of each other for acute or non-acute patients, scored 'yes' if tests occurred in excess of 48 hours for non-acute patients, scored 'no' if tests occurred in excess of 48 hours for acute patients. An acute patient defined as a participant diagnosed with stroke (or other neurological condition) within 6 weeks of the test.

Three studies performed the VFSS within 24 hours of the CSA^{61, 93, 94} and one study performed the VFSS within 48 hours⁵⁹ and one study⁹⁶ performed one test immediately before or after the other test, this is considered a reasonable time period. One study performed the VFSS within 96 hours of the CSA⁹⁰ which was not considered a reasonable time period as the study population included acute stroke patients –

diagnosis of stroke within 35 days of the assessment. Another study⁹⁵ completed the CSA within 0-16 days of referral and the VFSS within 0 – 47 days. This was also considered an unreasonable time period as the study included the population of acute stroke patients.

Two studies provided insufficient information to determine whether or not the time period between tests was reasonable.^{91, 98} One study⁹⁸ stated the VFSS was performed within one week of the CSA, however the study did not state if the patients receiving the test were recently diagnosed with dysphagia or long-term dysphagia patients. The second study⁹¹ stated the tests occurred within 72 hours however it is also unclear in this study the time post onset of dysphagia.

Four studies did not state the time interval between the tests.^{89, 92, 97, 99}

5. Did the whole sample or a random selection of the sample receive verification using the reference standard?

In one study⁸⁹ this process was unclear as there were 249 eligible participants, however only results for 238 participants were reported. In the remaining studies the whole sample received verification using the reference standard.

6. Did participants receive the same reference standard regardless of the index test result?

In the majority of studies all patients received the same reference standard regardless of the index test result. However in one study⁹⁶ five of the 96 participants received fiberoptic endoscopic evaluation of swallow (FEES) as the reference standard and in another study⁸⁹ this information was unclear as it was not explicitly stated.

7. Was the reference standard independent of the index test?

In all studies the index test did not form part of the reference standard and the tests were conducted independently.

8. Was the execution of the index test described in sufficient detail to permit its replication?

This item scored 'yes' if the authors outlined the process followed during the index test including: setting, position of the patient and consistencies used during the swallow trial (e.g. 5mls thin fluid, 10mls thin fluid). Ten of the papers described the index test in sufficient detail to permit its replication.

One paper ⁸⁹ provided the protocol used for the index test as an appendix item however the authors did not detail how this protocol was implemented or the oral trials the patient was given during the test. Another paper ⁸⁸ also outlined the process for the clinical swallow assessment such as the areas assessed but did not specify how the clinician executed the assessment or the oral trials used during the test.

9. Was the execution of the reference standard described in sufficient detail to permit its replication?

This item scored 'yes' if the authors outlined the process followed during the reference test including: setting, position of the patient and consistencies used during the swallow trial (e.g. 5mls thin fluid, 10mls thin fluid). Nine out of the 13 included studies described the reference standard in sufficient detail to permit its replication.

Two papers ^{88, 89} cited a standardised protocol in the reference list but did not provide an explanation in the paper of how the test was implemented. For example the papers did not explain the type of food and fluid offered to the patient, or the order in which the patient received the food/fluid trials.

One study⁸⁸ provided some information on the equipment used during the reference test and the consistencies trialled but did not explain the amount given or the decision making process used to determine which consistency each patient was offered.

One paper⁹⁴ cited an adapted protocol but only the reference for the original protocol was provided. Insufficient information was provided to be able to replicate the reference test.

One paper⁹⁸ clearly stated the consistencies offered but did not describe the contrast or recipes used to make the various viscosities, the positioning of the patient or the order for presenting each consistency to the patient.

10. Were the index test results interpreted without knowledge of the results of the reference standard?

The majority of papers explicitly stated that the clinician scoring the index test was 'blind' to the results of the reference standard or the index test was scored prior to the reference standard test occurring which implies the results of the reference test were not known. Two papers^{94, 97} did not state whether or not the index test results were interpreted without knowledge of the results of the reference standard and were rated as 'unclear'.

11. Were the reference standard results interpreted without knowledge of the results of the index test?

The majority of papers explicitly stated that the clinician scoring the reference test was 'blind' to the results of the index test. Two papers^{89, 97} did not state whether or not the index test results were interpreted without knowledge of the results of the reference standard and were rated as 'unclear'.

12. Were the same clinical data available when the test results were interpreted as would be available when used in clinical practice?

All studies used clinical data which would have been available when the test was used in clinical practice such as medical records, medical charts and time post onset of symptoms.

13. Were uninterpretable, indeterminate or intermediate test results reported?

The majority of papers scored 'n/a' for this question as all participants were evaluated using the index test and reference standard and results for each study were reported.

One study ⁹⁴ scored 'yes' for this as the paper reported 3 participants results that were in this category: one x-ray film was accidentally erased and two x-ray films were technically poor and could not be reported.

One study ⁸⁹ scored 'unclear' for this as 249 patients were initially recruited however only 238 results were reported for the reference standard test. It is unclear if the remaining 11 participants withdrew or if their test results were uninterpretable or indeterminate.

14. Were withdrawals from the study explained?

The majority of papers scored 'n/a' for this questions as results for all initial participants were provided.

One paper ⁹⁴ explained that a patient withdrew as they were too drowsy to participate in the index test assessment. One study ⁸⁹ scored 'unclear' for this as 249 patients were initially recruited however only 238 results were reported for the reference standard test. It is unclear if the remaining 11 participants withdrew or if their test results were uninterpretable or indeterminate.

4.1.4 Assessing methodological quality using the STARD checklist

In this systematic review the STARD checklist was used to extract methodological quality data in addition to those of the QUADAS checklist. The data extracted for each individual item is presented below (Q1- Q25), the full checklist is presented in Appendix V. In addition to the QUADAS checklist items, the STARD checklist was used to assess further aspects of methodological quality as well as in data extraction.

Performance of included studies against individual STARD Checklist Items

1. Was the study identified as a study of diagnostic test accuracy?

The majority of the included studies included the terms 'sensitivity and specificity' in the abstract or title.^{61, 90, 92, 94, 96, 97} One study⁹⁵ included the term 'diagnostic accuracy' in the title. Four studies^{59, 89, 91, 98} did not include either of these terms but did include the terms 'predictive value'.

2. Research questions or study aims stated such as, estimating diagnostic accuracy or comparing accuracy between tests.

The study aims of all included studies included a comparison of the diagnostic test accuracy of the index test - clinical swallow assessment with the reference standard - video fluoroscopic swallow study, or the equivalent titles for index test and reference test.

3. Description of study population including setting, inclusion and exclusion criteria.

All papers stated the participants were referred for feeding and/or swallow assessment. The majority of papers described in sufficient detail the study population and setting.^{59, 90-92, 94-98} The inclusion and exclusion criteria was generally not well reported. Specific details of these findings are

discussed below. One study ⁸⁹ provided limited information regarding study population and setting and stated that 95% of the participants were adults with a neurological cause for dysphagia.

The most common setting was a hospital setting ^{59, 90, 92, 94, 95, 98} followed by a rehabilitation centre/unit ^{91, 97} and one study was set in a medical centre. ⁹⁶ Three studies did not specify setting. ^{61, 88, 93}

Several studies did not explicitly state any exclusion criteria. ^{59, 91, 97, 98} Three papers excluded patients with a history of head and neck cancer; ^{90, 92, 96} a tracheostomy ^{61, 90, 93} or previous history of dysphagia. ^{61, 93, 95} Two papers excluded patients with a structural anomaly. ^{61, 93} Remaining exclusion criteria included: presence of a brain tumour ⁹⁶, neurosurgery ⁹⁶, gastrostomy feeds ⁹⁰, progressive neurological disease ⁹⁰ and patient not providing consent. ⁹⁴ Patient consent may have been implied in the other included studies.

4. Participant recruitment: was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?

All included studies provided this information. The majority of patients were recruited based on their diagnosis of a recent stroke ^{61, 93-96} and the remaining five studies recruited patients based on their referral for a feeding/swallow assessment due to signs and symptoms consistent with dysphagia. ^{59, 88, 90, 91, 97}

5. Participant sampling: was the study population a consecutive series of participants defined by the selection criteria in item 3 and 4? If not, specify how participants were further selected.

The majority of included studies explicitly stated that the participants were sampled from a consecutive series based on the pre-defined criteria. ^{59, 61, 91-93, 95-97} Several studies did not specify this information ^{88, 89, 98} and one

study ⁹⁰ stated 'purposive sampling' was used to identify the patients within the population who met specific pre-determined criteria.

6. Data collection: was data collection planned before the index test and reference standard were performed (prospective) or after (retrospective study)?

The majority of included studies used a prospective study design. Study design for three studies was not clear. ^{97, 99, 100}

7. The reference standard and its rationale.

All studies used video fluoroscopic swallow study (or synonymous study name such as Modified Barium Swallow) as the reference standard. This is well documented previously in this thesis as being an acceptable gold standard assessment for oropharyngeal aspiration.

8. Technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard.

The majority of included studies provided sufficient details outlining the method for the index test and reference test including: sequence of examination, positioning of the patient, oral trials offered and the scoring/assessment process used. ^{59, 88, 90, 92-98} Several studies cited a reference or protocol used for one or both of the tests but did not provided any further details. ^{61, 89, 91}

9. Definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.

All included studies provided a detailed description of the categories and/or cut-offs used by the clinician to determine the presence or absence

of oropharyngeal aspiration. All included studies provided these details for the index test and the reference test.

10. The number, training and expertise of the persons executing and reading the index tests and the reference standard.

Details varied between studies and within studies with regard to the clinician/therapist performing each test. Three studies specified that the clinician performing the index and reference test were specifically trained in dysphagia,^{59, 93, 98} in one study this was only specified for the index test⁹⁴ and information regarding the reference test was not specified. Two studies specified the index test was performed by a certified speech language pathologist.^{89, 91} Linden et. al did not provide details for the reference test and Splaingard et al. stated this was performed by a physician and a speech language pathologist. For several studies there was paucity in this information.^{61, 90, 92, 95-97, 99}

11. Were the readers of the index test and reference standard blind to the results of the other test?

The majority of studies explicitly stated the reader of the test in question was blind to the results of the other test.^{59, 61, 88, 90-96} This information was not specified in three papers.^{89, 97, 98}

12. Methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods use to quantify uncertainty (e.g. 95% confidence intervals). The majority of studies explicitly stated this information^{59, 61, 92-97} most frequently the reporting of 95% confidence intervals. Four studies did not explicitly state this information.^{89-91, 98}

13. Methods for calculating test reproducibility, if done.

None of the included studies described calculation of test reproducibility.

14. When study was performed including beginning and end dates of recruitment.

Approximately half of the included studies did not include this information.^{59, 61, 89, 92, 93, 97, 98} Of the remaining six studies, three stated the beginning and end dates of recruitment^{88, 90, 96} and three stated the time period only.^{91, 94, 95} The time period ranged from four months to two years.

15. Clinical and demographic characteristics of the study population.

All studies provided basic information such as: age, gender and medical diagnosis. The majority of studies described patients as presenting with dysphagic symptoms.

16. The number of participants satisfying the criteria for inclusion who did or did not undergo the index tests and/or reference standard.

Details were not explicitly stated in any of the included studies. All participants underwent both tests. This information was ambiguous in one study⁸⁹ where there were 249 patients described in the methods section however only results for 238 patients were provided.

17. Time interval between the index test and reference standard and any treatment administered in between.

The time interval between the index test and reference standard did vary between studies. The possible implications of this are further discussed in the results section. The table of included studies Appendix VII states this information for each of the included studies.

No studies described the prescription of any treatment, therapy or intervention between administration of the index test and the reference standard and it is assumed that there was no treatment administered in this time.

18. Distribution of severity of disease in those with the target condition.

None of the included studies described the severity of aspiration, studies reported if aspiration was present or absent on VFSS.

19. A cross tabulation of the results of the index test by the results of the reference standard.

Four of the included studies provided this data in the traditional 2 x 2 table.^{59, 90, 93, 95} The remaining studies presented this information as a percentage, in narrative form or only provided the overall sensitivity and specificity. In these instances, the raw data was manually extracted from the provided summarised data.

20. Any adverse events from performing the index test or the reference standard.

No studies reported any adverse events during the conduct of either the index test or the reference standard.

21. Estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals).

There was variable reporting of this information between studies. Several studies reported sensitivity and specificity values and associated p-values but without confidence intervals.^{61, 93, 94, 97, 98} Three studies^{88, 89, 91} reported the raw data, for example, number of true positive and false positive results without confidence intervals, two studies reported sensitivity and specificity results with confidence intervals^{92, 96} and only two studies^{59, 95} reported sensitivity and specificity, positive and negative predictive values and confidence intervals. One study⁹⁰ presented sensitivity and specificity, positive and negative predictive values and positive and negative likelihood ratios.

22. How indeterminate results, missing data and outliers of the index test were handled.

This was not applicable for nearly all included studies as all patient results were reported. The exception was ¹⁰⁰ that initially described 249 patients in the methods section of the study however only results for 238 patients were presented without comment or explanation regarding the missing data of 11 patients.

23. Estimates of variability of diagnostic accuracy between subgroups of participants, readers or centres, if done.

This information was not reported in any of the included studies.

24. Estimates of test reproducibility, if done.

This information was not reported in any of the included studies.

25. Discuss the clinical applicability of the study findings.

All of the included studies described the clinical applicability of the study findings regarding the diagnostic accuracy of clinical swallow assessment for detecting aspiration.

4.1.5 Data from individual studies

The following section presents the salient features and important characteristics of each of the studies included in this systematic review. One author ⁹⁸ was contacted and provided the data for the number of patients who did aspirate and the number of patients who did not aspirate in the study.

Barbiera et al. 2006:

This study was conducted in Italy and included 47 patients within a neuro-motor rehabilitation program. Patients were aged 16 – 80 years at recruitment. Diagnoses included: cerebrovascular accident (n=30), head-brain trauma (n=7), brain tumour (n=4), Parkinson's disease (n=2), amyotrophic lateral sclerosis (n=2), multiple sclerosis (n=1), olivopontocerebellar atrophy (n=1). The study aim included demonstrating the importance of video fluorography swallow study (VFSS). All patients underwent a speech assessment and a VFSS and assessed for presence or absence of aspiration. A Radiologist performed the VFSS however the study did not explicitly state who performed the speech assessment. Time interval between the index and reference tests were not stated.

The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Briefly, the inclusion criteria was: patients presenting with oropharyngeal dysphagia and able to undergo the speech assessment and VFSS. Details of the speech assessment reported as: assessment of history and physical examination and speech assessment including swallowing trials. The type and/or amount of food and/or fluid trials was not specified. Patients were scored as either: dysphagic with suspected aspiration, dysphagic with no signs of aspiration, non-dysphagic. Details of VFSS reported as: patients seated in a wheel chair and trialled with at least two different consistencies of barium and rated as presence or absence of aspiration.

Patient results were presented in narrative form and included the number of patients scored as 'aspirating' based on the speech assessment and number of patients confirmed aspirating based on the VFSS. Presented results were used to populate a 2 x 2 table and calculate sensitivity and specificity.

Calculated results:

	VFSS (+)	VFSS (-)
Speech assessment (+)	17	12
Speech assessment (-)	4	14

Sensitivity: 81%, Specificity: 54%

Baylow et al. 2009:

This study was conducted in USA and included 15 acute stroke patients. Patients were aged 48 – 80 years with a diagnosis of CVA. The mean number of days post-onset of CVA was 9.2 (1 – 35). The majority (n=11) within 2 weeks of stroke. Although not the primary study aim, the correlation between clinical and video fluoroscopic findings was examined. Purposive sampling (non probability sampling) was used to identify patients within the population who met specific criteria including referral by a physician for the assessment of a possible swallowing disorder. All participants underwent a CSA and VFSS with a range from 1 to 35 days of admission. CSA was performed by a speech pathologist using the Northwestern Dysphagia Patient Check Sheet – a screening tool comprised of 28 patient variables.⁹⁰ Two trial swallows per fluid or food consistency were offered. A Radiologist performed the VFSS with the principal investigator (primary author of study) within 96 hours after completion of CSA.

The primary aim of this study was to examine the sensitivity and specificity of the accuracy of using the chin down posture during the CSA. However, results of CSA compared with VFSS were presented when patient used chin down posture and usual positioning. For the purposes of this systematic review, only results regarding use of usual head/neck position were extracted. The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Briefly, inclusion criteria reported as: diagnosis of acute CVA provided by a medical doctor, referral by physician for assessment of possible swallowing problem, adequate consciousness to participate in swallowing

evaluation, medical stability and no history progressive neurologic disease, head and neck cancer, dysphagia and/or presence of a tracheostomy or gastrostomy tube. Details of the clinical swallow assessment reported as: two trials of each consistency were performed – 1cc thin liquid, 1cc pudding, ¼ piece of cookie. A dichotomous scoring system was used to rate each variable as either safe or unsafe. The same number and type of bolus trials used for the CSA were also used for the VFSS. VFSS films were reviewed by a speech pathologist and radiologist for presence or absence of aspiration.

Patient results were presented in table form including – sensitivity and specificity, as a percentage. Presented results were used to populate a 2 x 2 table. The total number of results is 30 to account for two swallows per patient.

Calculated results:

	VFSS (+)	VFSS (-)
Speech assessment (+)	2	5
Speech assessment (-)	3	20

Sensitivity: 40%, Specificity: 80%

DeMatteo et al. 2005:

This prospective study was conducted in Canada and included 59 infants and children presenting with feeding and/or swallowing difficulties. 62% of the children were younger than 12 months. The aim of the study was to evaluate the accuracy of the clinical evaluation compared with VFSS in the detection of penetration and aspiration in children. Participants first underwent the clinical swallow assessment then the VFSS. VFSS was performed on the same day or within 48 hours of the clinical swallow assessment. The table of included studies (Appendix VII) provides a summary of the important characteristics of each included study.

The inclusion criteria was reported as: consecutive recruitment over a 15 month period including inpatients and outpatients with any diagnosis and

participant aged 0 – 15 years. The clinical swallow assessment included an oral motor examination and swallowing evaluation. The child was fed by his/her usual care giver in the usual feeding position using their own foods and feeding utensils. During the video fluoroscopic swallow study, the child was positioned in either their own seating or as close as position as possible using radiology chairs. Pre mixed liquid barium consistencies were used for the children and non-ionic x-ray contrast solutions were used for the infant feeds. Each child was fed by his/her typical caregiver or the Occupational Therapist. A clinician different from the clinical swallow evaluation completed the VFSS evaluation in consultation with the Radiologist.

Patient results were categorised using the 2 x 2 table designations and sensitivity and specificity results were presented as percentages.

Presented results:

	MBS (+)	MBS (-)
Clinical swallow evaluation (+)	22	19
Clinical swallow evaluation (-)	2	16

Sensitivity: 92%, Specificity 46%.

DePippo et al. 1992:

This cross sectional study was conducted in the USA and included 44 adult patients on a stroke rehabilitation unit with a confirmed diagnosis of stroke. The mean age of the patients was 71years. The aim of the study was to determine the sensitivity and specificity of a water swallowing test to predict aspiration using Modified Barium Swallow (MBS) as the reference test. The study did not explicitly state who performed the water swallowing test. The Modified Barium swallow studies were reviewed by two speech pathologists who determined the presence or absence of aspiration. The time interval between the index and reference test was not

specified. The table of included studies (Appendix VII) provides a summary of the important characteristics of each included study.

Briefly, the inclusion criteria was reported as patients presenting with one or more of the following features suggestive of dysphagia: bilateral hemispheric stroke, brain-stem stroke, history of pneumonia during acute stroke phase, coughing associated with feeding, failure to consume half of meals, prolonged time required for feeding and non-oral feeding program in progress. During the water swallow test, patients were given 3 oz. (89mls) of water and asked to drink consecutively from a cup. The finding of coughing during or for 1 minute after completions or post swallow wet-hoarse voice scored as 'abnormal'. During the modified barium swallow, patients were given 5 mls of thin barium liquid, 5 mls thick barium liquid, 5 mls barium pudding, one quarter barium cookie, 20mLs thin barium liquid (taken in one swallow) and 30 mls thin barium liquid (consecutive swallows). Patients were seated in an upright position. The films were recorded and reviewed by two speech pathologists who determined presence or absence of aspiration.

Patient results were not categorised using the 2 x 2 table designations. Results were presented in two columns and included: total number of patients, total number with aspiration on MBS, total number with abnormal water swallow test, total with abnormal water swallow test and aspiration on MBS, total false-positive results, total false-negative results, sensitivity and specificity. This data was then categorised into a 2 x 2 table. It is interesting to note the calculated sensitivity and specificity did not match study data. Below are the results calculated for this systematic review and secondly, results provided in the paper.

Results calculated for this systematic review

	MBS (+)	MBS (-)
Water swallow test (+)	16	11
Water swallow test (-)	4	13

Sensitivity: 80%, Specificity 54%.

Results presented in the paper:

Total number of patients	44
Total with aspiration on MBS	20

Sensitivity: 76%, Specificity: 59%

Hammond et al. 2009

This study was conducted in the USA and included 96 adult patients with a mean age of 68 years, following a recent ischemic stroke admitted to Veteran Affairs Medical Centre. The time post onset of symptoms or stroke diagnosis not provided. Authors of this study hypothesised that objective measures of voluntary cough would improve the accuracy of the clinical evaluation of swallow to predict those patients who are at risk.

In this study, two reference tests were available – VFSS or FEES (fiberoptic endoscopic evaluation of swallow). 91 patients underwent VFSS as the reference test, 5 patients underwent FEES. Study authors were contacted to retrieve data for VFSS patients only but the authors did not respond to the request. Therefore the study results need to be interpreted with the knowledge that 5 patients underwent a FEES (not a VFSS) to confirm the presence or absence of aspiration. The CSA was performed by speech pathologists and the films of the VFSS or FEES were analysed and rated for presence or absence of aspiration by a speech pathologist using standard criteria. Briefly, a FEES involves a thin endoscope passing through the person's nare as they are given food and fluid to swallow. The bolus is dyed so it can be visualised by the person performing the study. The images are often recorded on a screen/monitor and reviewed by the clinicians performing the test.

The participant inclusion criteria was reported as: consecutive, consenting patients with recent ischemic CVA. The exclusion criteria was reported as: patients with a history of radiation therapy to the head and neck, brain tumour or brain surgery excluded. The CSA included an assessment of reflexive cough; coughing or choking after ice chips or water and/or an absent swallow. Patients were classified as aspirators based on the

clinical assessment if any of the three assessments were judged as abnormal. During the VFSS or FEES, the patient was seated upright and the tests were videotaped for later analysis. It was unclear who performed the tests, however films were reviewed and rated by a speech pathologist blind to results of the CSA. The index test was performed immediately before or after the index test.

The sensitivity and specificity of CSA was presented as a percentage. The raw data was extracted to populate the 2 x 2 table:

Calculated results:

	VFSS (+)	VFSS (-)
Bedside swallow assessment (+)	19	11
Bedside swallow assessment (-)	14	52

Sensitivity: 58%

Specificity: 83%

Linden et al. 1993

This study was conducted in the USA and although the study initially states 249 patients participated, only results for 238 patients presented. Exact details of the study setting were not specified. Authors report approximately 95% of patients were adults with a neurological cause for their dysphagia, with the most common being stroke. Underlying causes for the remaining 5% were not reported. It is unclear if the population included patients with a tracheostomy, which was an exclusion criteria of this systematic review. The term 'subglottic penetration' is used in this paper, the definition of which is equivalent to 'aspiration' as defined in this thesis. The aim of this study was to determine which clinical indicators from the Dysarthria/Dysphagia Battery were predictive of subglottic penetration as documented by video fluorographic swallowing studies (VFSS). The Dysarthria/Dysphagia Battery was the chosen checklist/CSA used by the clinicians in this study and is comparable to the CSAs used in the other included studies. The CSA was performed by a speech-language pathologist 'knowledgeable' about dysphagia however the details for VFSS were not stated. The time interval between clinical

swallow assessment and VFSS not stated either. The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Selection process and inclusion criteria not specified. References and protocols cited for the clinical swallow assessment and video fluorographic swallowing studies, however details of the assessments such as patient positioning and swallowing trials not provided. Clinical swallow assessment scored using a binary system – clinician rated the patient as either showing signs of subglottic penetration (aspiration) or no signs/evidence of subglottic penetration. Details for scoring of VFSS are not stated. Patient results are presented in a 2 x 2 table. Based on this data, sensitivity and specificity have been calculated for the 238 patients. The impact of the remaining patients of the sensitivity and specificity of CSA is unknown.

Data extracted:

	VFSS (+)	VFSS (-)
Clinical swallow assessment (+)	64	46
Clinical swallow assessment (-)	36	92

Sensitivity: 64%, Specificity: 66%

Mann et al. 2000

This prospective study was conducted in Australia and included 128 adult patients presenting to the hospital's acute stroke unit with first-ever acute stroke. The mean age of the sample was 71 years of age. Patients were excluded if they had a previous swallowing impairment or medical condition that could affect swallowing function. The study aim was to determine, in patients with acute first-ever stroke the accuracy of the standardised bedside clinical assessment compared with video fluoroscopy. Two speech pathologists blinded to the video fluoroscopic findings independently assessed swallowing function at bedside using standardised methods provided as an appendix item in the paper.

The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Details of bedside swallow assessment: history was taken from patient and/or relatives, an oral examination and a swallow trial of 5 ml of water, 20 ml of water and thickened fluid (if appropriate). Details of video fluoroscopy: patients positioned upright and swallow consecutive boluses (5 ml, 10 ml) of thin-liquid then thick-liquid. Video fluoroscopic examinations were evaluated blind to results of the bedside swallow assessment.

Patient results are presented in 2 x 2 table and the sensitivity and specificity presented as a percentage.

Presented results:

	VFSS (+)	VFSS (-)
Bedside swallow assessment (+)	26	37
Bedside swallow assessment (-)	2	63

Sensitivity: 93%, Specificity: 63%

McCullough et al. 2001

This study was conducted in the USA and included 60 adult patients with recent thromboembolic stroke. Mean number of days post onset of CVA was 5.98 days with a range of 1 – 42 days. Patients were excluded from the study if they had a structural anomaly, presence or recent history of tracheostomy, dysphagia prior to the stroke. The study aim was to investigate the sensitivity and specificity of clinical/bedside examination signs for predicting aspiration on video fluoroscopic examination of swallowing. It was not stated who performed the tests.

The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Details of bedside swallow assessment: examination of oral structures and reflexes, voice measures and trial swallows using 5 cc bolus size of thin and thickened liquids and ¼ cookie for a solid. Details of video

fluoroscopy: thin and thickened liquids were given to patient in 5cc bolus size. Water and barium and juice and barium were used as the fluid recipes. Patients were categorised as either having aspiration present or absent.

Patient results were presented in percentages for sensitivity and specificity in the paper. The data was used to populate a 2 x 2 table.

Calculated results:

	VFSS (+)	VFSS (-)
Bedside swallow assessment (+)	17	14
Bedside swallow assessment (-)	5	24

Sensitivity: 77%, Specificity: 63%

McCullough et al. 2005

This study was conducted in the USA and included 165 consecutively enrolled adult patients, recruited from Veteran's Affairs Medical Centres. All participants were diagnosed with a stroke within 6 weeks of the time of examination. The study aim included measuring the accuracy of clinical swallowing examination (CSE) or combinations of CSE to detect aspiration in stroke patients using Video Fluoroscopic Swallow Examination (VFSE) as the reference test. The CSE was performed by a speech pathologist and the VFSE was performed by another speech pathologist blind to the results of the CSE. All clinicians had over 200 hours experience with CSE and VFSE examinations of swallowing. The VFSE was performed within 24 hours after completion of the CSE. The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Briefly, the inclusion criteria was reported as: occurrence of a stroke within 6 weeks of the time of examination. The exclusion criteria included: presence of a structural anomaly that could affect swallowing, presence or recent history of tracheostomy, reported history of dysphagia prior to

stroke. The clinical swallowing examination was comprised of four sections: history, oral motor, voice and speech praxis and trial swallows. The trial swallows included two swallows of each consistency: 5 ml thin liquid, 10 ml thin liquid, thick liquid, puree and solid (1/4 cookie), liquids from a pill cup, puree and solids from a spoon. The clinician judged whether or not it was 'safe' to proceed with the 3 oz. water swallow test. Clinicians used a binary rating system and recorded presence or absence of aspiration. During the video fluoroscopic swallow examination participants were seated upright in a wheel chair or stretcher chair. The patients were given two 5 ml then two 10-ml thin liquid swallows (50/50 mixture of water and Barium); two 5 mls of thickened liquid; two 5 ml of applesauce; two solids (cookie). Finally, consecutive swallowing 3 oz. of thin liquid barium was given to the patient. The films were recorded and reviewed and rated as aspiration present or absent in the patient.

The patient results were presented in table form and included: sensitivity, specificity, positive predictive value, negative predictive value and positive likelihood ratio. The data provided was extracted and entered into the online diagnostic calculator, ¹⁰¹ to populate a 2 x 2 table.

Calculated results:

	VFSE (+)	VFSE (-)
Clinical swallowing examination (+)	23	13
Clinical swallowing examination (-)	20	109

Sensitivity: 54%, Specificity: 89%

Smithard et al. 1998

This study was conducted in the UK and included 83 patients in a hospital setting assessed by one of the possible two speech pathologists (SLT1). Only results for SLT1 are presented in this systematic review as the data for SLT2 were not explicitly reported in the study. The details of patient withdrawals and exclusions were clearly described in the paper. The study

population included adults with an acute onset of stroke. 'Acute onset' was defined as presentation to the hospital within 24 hours of symptoms. Patients were excluded if they presented after 24 hours or if there was the presence of a serious intercurrent illness (e.g. advanced malignancy). The study aim was to investigate the ability of bedside swallowing assessment to exclude aspiration following acute stroke.

The bedside swallow assessment was performed by a speech language therapist trained in the management of dysphagia. The therapist performing the Video fluoroscopy was not specified. The bedside swallow assessment was performed within 24 hours of the video fluoroscopy. This study also considered the difference in assessment carried out by different team members therefore patients were assessed by two doctors (Doc1 and Doc2) and two speech-language therapists (SLT1 and SLT2). As mentioned previously, only results for SLT1 are presented in this systematic review as the data for SLT2 were not explicitly reported in the study.

The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Briefly, the bedside swallow assessment was reported as a 'standardised bedside swallowing assessment', no further details were provided. The SLT recorded if the patient's swallow was safe or unsafe. During the video fluoroscopy, patients were offered different consistencies and volumes of barium using a standard protocol. The results of the VF were reported as aspirating or not aspirating.

Patient results presented in narrative form as well as a table with the following information: number patients (n=83), sensitivity, specificity, positive and negative predictive value. Data used to populate a 2 x 2 table.

Calculated results:

	VFSS (+)	VFSS (-)
Bedside swallow assessment (+)	9	9
Bedside swallow assessment (-)	10	55

Sensitivity: 47%, Specificity: 86%

Splaingard et al. 1988

This study was conducted in the US and included 107 patients in a comprehensive free-standing rehabilitation centre. Study population: 87 adult stroke patients, 16 adult brain injury patients, ten children (less than 15 years) and four adults with chronic neuromuscular diseases. Twenty three participants had either a tracheostomy, a gastrostomy or a nasogastric tube, or a combination of two of these. Results were provided for the 84 patients without an 'appliance'. These are the results presented in this systematic review as presence of a tracheostomy was an exclusion criteria for this systematic review. Interestingly, when the results were calculated using the online diagnostic calculator,¹⁰¹ there was a total of 85 patients. The reason for this discrepancy is unclear. The following data was entered into the on-line calculator as presented in the paper: prevalence, sensitivity, specificity and total sample size.

The study aim was to evaluate information regarding aspiration obtained from video fluoroscopy versus bedside clinical assessment. The bedside clinical assessment was performed by certified speech-language pathologists (SLP). The number of SLPs was not stated. The video fluoroscopy was performed by a physician and SLP blinded to the results of the bedside clinical swallow assessment. The video fluoroscopy occurred within 72 hours of the clinical swallow evaluation. The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

This study included patients referred by physicians for evaluation of possible swallowing dysfunction. Clinical swallow evaluation details included: a chart and case history review; oral examination; swallowing trials – liquids (juice, nectar, frosty), puree food, ground meat and solids. The SLP completed a standardised bedside clinical assessment form and classified the patient using a 5 point scale. A score of 4 or 5 indicated the presence of aspiration, whereas 1 or 2 indicated not aspiration. The video fluoroscopic swallow study was reported to include a 'standardised' technique with the patients in an upright position. The following sequence of food and fluid was reportedly used during the VFSS and mixed together

with barium: pudding, extra thick liquid, thick liquid, thin liquids and solids. The presence or absence of aspiration was noted with each consistency.

Patient results were presented in narrative form and included: number of patients scored as 'aspirating' on video fluoroscopy (n=29) and the number detected by clinical swallow evaluation (n=6). The presented results were used to populate a 2 x 2 table and calculate sensitivity and specificity.

Calculated results:

	VFSS (+)	VFSS (-)
Speech assessment (+)	6	4
Speech assessment (-)	23	52

Sensitivity: 20%, Specificity: 90%

Tohara et al. 2003

This study was conducted in Japan and included 63 patients with signs or symptoms of dysphagia. Patients ranged in age from 18 – 83 years, with an average age of 63 years. Diagnoses included: cerebrovascular accident (n=36), neuropathy or myopathy (n=7), traumatic encephalopathy (n=6), respiratory disease (n=5), brain tumour (n=4), oral or pharyngeal tumours (n=2), cervical spinal cord injury (n=1), inanition (n=1), unknown aetiology (n=1). The study aim was to determine whether 3 non video fluorographic (non-VFG) bedside examinations would constitute an effective screening battery for aspiration. All patients underwent both the clinical swallow assessment and a VFSS and were assessed for presence or absence of aspiration. The bedside swallow assessment and VFSS was performed by a rehabilitation physician and/or dentist trained in dysphagia. This was the only study where the VFSS was in the form of a pre swallow x-ray and a post swallow x-ray, rather than a motion x-ray. Time interval between these tests not stated.

The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Patient results presented as sensitivity and specificity – percentage. Author contacted for raw data, this information was provided. Presented results used to populate a 2 x 2 table and calculate sensitivity and specificity.

Calculated results:

	VFSS (+)	VFSS (-)
Speech assessment (+)	20	4
Speech assessment (-)	9	30

Sensitivity: 69%, Specificity: 88%

The original paper also reported a specificity of 88%, however sensitivity was reported as 70%.

Zhou et al. 2011:

This study was conducted in France and included 107 patients following diagnosis of stroke. Mean time after stroke was 6.5+/-5.2 months. Patients aged 67.8+/-13.3 years. This formed phase 2 of a two part study. The first phase looked at the establishment of an efficient screening tool to detect aspiration and the second phase addressed validating this tool. All patients underwent clinical swallow assessment using the tool title the ‘Practical Aspiration Screening Scheme’. ‘This involved six features of clinical swallow assessment – absence of archaic reflexes, presence of velar reflex, voluntary swallowing, absence of dysphonia, presence of gag reflex and voluntary glottic closure.’⁹² Each item was allocated a number of points. For example, voluntary swallowing was assigned 7 points and presence of a gag reflex was assigned 6 points. This is a validated test developed by the authors of the paper. Patients scoring between 14-28 were considered ‘unclear’ and also went on to have a 3 oz. water swallow test. Based on this information the patient was scored as either aspiration present or absent. Two rehabilitation physicians specialised in the examination performed the VFSS. The time interval between these tests was not stated for phase 2 of the study.

The table of included studies (Appendix VII) provides a summary of the study details as reported by the authors.

Inclusion criteria: patients with CVA, diagnosed objective lesions on cerebral imaging. Exclusion criteria: sub-arachnoid haemorrhage, transient ischaemic attacks, head and neck cancer, below 18 years of age.

Patient results were presented in narrative form and included: percentage of patients aspirating and percentage detected by CSA. Presented results used to populate a 2 x 2 table and calculate sensitivity and specificity.

Calculated results:

	VFSS (+)	VFSS (-)
Speech assessment (+)	48	10
Speech assessment (-)	6	43

Sensitivity: 89%, Specificity: 81%

4.1.6 Data analysis and synthesis

Meta-analysis allows the data of sufficiently similar studies to be summarised and in this case, a summary estimate of the clinical swallow assessment test accuracy, to be calculated. Meta-analysis in studies of DTA is challenging due to the fact that there are two facets to diagnostic performance, for example, sensitivity and specificity that need separate estimation, rather than a single measure such as odds ratio or relative risk.⁸⁶ One of the results of this study is an exploration of the most appropriate method for identifying the presence of heterogeneity and threshold effects in a dataset of DTA study results, and in statistically combining them in meta-analysis.

The following data was extracted from individual studies: true positive, false positive, false negative, true negative (Appendix VI) and combined in meta-analysis. The majority of studies did not provide this data in the traditional 2 x 2 table, therefore they were calculated manually with the provided sensitivities, specificities and prevalence using a diagnostic

calculator.¹⁰¹ The results of TP, FP, TN and FN for each included study as well as sensitivity and specificity were plotted and presented as a forest plot using Revman 5.¹⁰² The forest plot shows two graphical sections – one depicting sensitivity and one depicting specificity, together with confidence intervals – and can be used to identify heterogeneity in the results of studies. Confidence intervals that do not overlap between studies indicates heterogeneous findings.⁸⁶ The paired results for sensitivity and specificity for each study were also plotted graphically as points in a summary receiver-operating characteristic (sROC) plot. The sROC plot highlights co-variation between sensitivity and specificity. In an sROC plot, the x-axis displays specificity obtained in studies in the review, the y-axis is the corresponding sensitivity. Estimation of a summary point shows the average sensitivity and specificity estimate of the study results with a confidence region surrounding this point. The point and confidence region can be estimated by using the bivariate random effects model.^{12, 13, 102} The bivariate method models the sensitivity and specificity directly and is regarded as having two levels corresponding to variation within and between studies.¹⁰² This model assumes real differences exist between study populations and procedures leading to errors that are not random.¹³ Estimates of the pooled sensitivity and specificity, the summary ROC plot and confidence regions were produced using the bivariate model proposed by Reitsma et al.¹⁰³ Neither JBI MAStARI or Cochrane Review Manager were able to perform meta-analysis or calculate summary estimates for diagnostic test accuracy studies therefore an external statistician computed the summarised data. The raw data from each included study was analysed using the 'metandi' command in Stata version 12.

The mean sensitivity and specificity estimate of the study results is presented graphically as a solid square on the plot. The sROC plot is presented subsequently in the next section of this paper and assists in the determination of the presence of threshold effects, which occur when there are variations between studies in the cut off value used to determine a positive or negative test result.⁸⁶

4.1.7 Meta-analysis

This systematic review has combined study results in meta-analysis and explicitly explored the sources of heterogeneity of study results. The sensitivity of CSA as compared with VFSS from the data extracted from the 13 included studies varied from 21% to 93%, the specificity from 46% to 93% (see Figure 4, Figure 5 and Table 3).

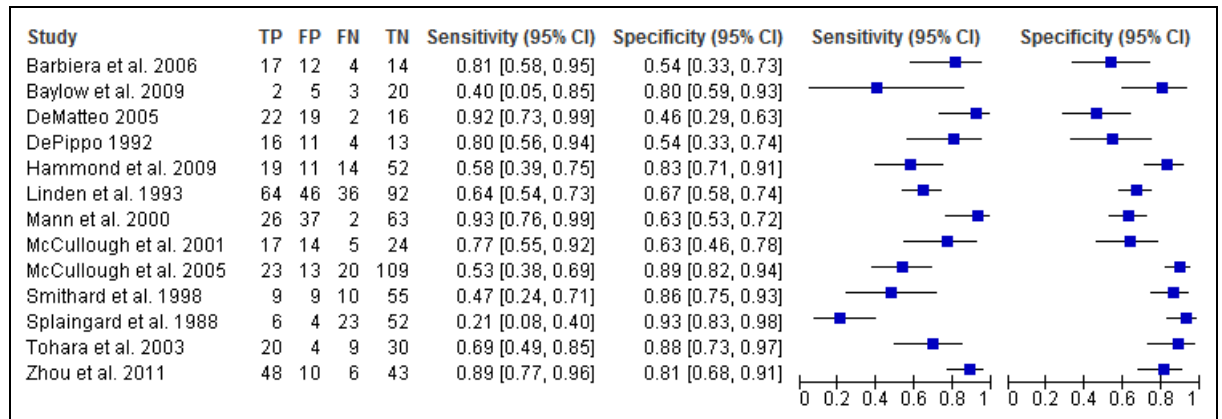


Figure 4 Forest plot of the included studies. TP = True Positive, FP = False Positive, FN = False Negative, TN = True Negative. Between brackets the 95% confidence intervals (CI) of sensitivity and specificity. The figure shows the estimated sensitivity and specificity of the study (blue/solid square) and its 95% confidence interval (black/solid horizontal line).

When reviewing forest plots for studies of DTA, homogeneity can be appraised by the amount of overlap between the confidence intervals.⁸⁶ This forest plot shows that apart from the Baylow et al study, which is very small and therefore has very wide confidence intervals, these results show a moderate amount of heterogeneity for sensitivity. For example, there are a number of studies^{91, 92, 98} where the confidence intervals do not overlap. There appears to be less heterogeneity with the specificity results (i.e. there are less studies that do not have overlapping confidence intervals). Overall, the forest plot shows a moderate amount of heterogeneity for the sensitivity results.

Table 3. Meta-analysis of diagnostic accuracy.

Log likelihood = -78.347976		Number of studies = 13			
	Coef.	Std. Err.	z	P> z	[95% Conf. Interval]
Bivariate					
E(logitSe)	.8769937	.3055828			.2780624
1.475925					
E(logitSp)	1.142858	.2341043			.6840216
1.601693					
Var(logitSe)	.976272	.4918483			.3636917
2.620646					
Var(logitSp)	.5838371	.2880929			.2219568
1.53573					
Corr(logits)	-.8563814	.1227582			-.974868
.360244					
HSROC					
Lambda	2.070824	.2184353			1.642698
2.498949					
Theta	-.2641947	.2744281			-.8020639
.2736745					
beta	-.2570597	.2606894	-0.99	0.324	-.7680015
.2538822					
s2alpha	.2168562	.1771486			.0437351
1.07526					
s2theta	.7007587	.3164629			.2891782
1.698132					
Summary pt.					
Se	.7061988	.0634029		.5690711	.8139563
Sp	.7582039	.0429185		.6646357	.8322549
DOR	7.537204	1.639144		4.921486	11.54315
LR+	2.920638	.3818314		2.260452	3.773637
LR-	.3874962	.0699074		.2720846	.5518627
1/LR-	2.58067	.4655736		1.812045	3.675328
Covariance between estimates of E(logitSe) & E(logitSp) -.0504454					

The results of the two-level logistic mixed effects model are presented in Table 3. The pooled sensitivity was 71% (57- 82 95% CI) and the pooled specificity was 76% (66-83 95% CI).

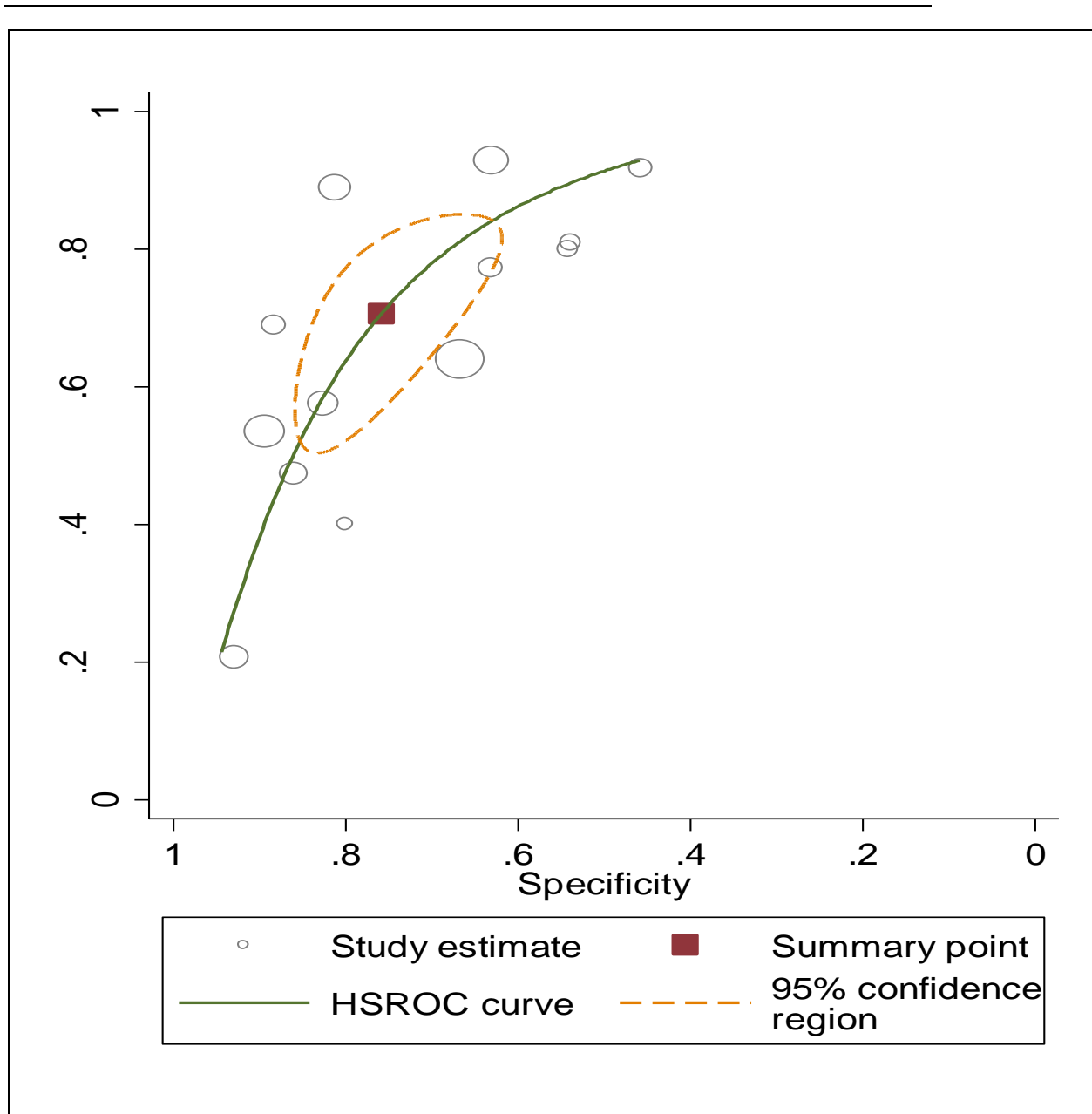


Figure 5 Summary receiver operator characteristic (sROC) plots of the sensitivity and specificity for all included studies and mean sensitivity and specificity estimate (red/solid) square of the study results and 95% CI (yellow) dashes.

In Figure 5 the x-axis (horizontal line) of the sROC plot displays the specificity obtained in the included studies in the systematic review and the y-axis (vertical line) shows the corresponding sensitivity. Each oval-shaped symbol represents the sensitivity-specificity point for each individual study. The greater the size of the oval the heavier the weight of the study in calculating the summary point. The mean sensitivity and specificity is 71% and 76% respectively and represented by the shaded square. The dashed line surrounding the square represents the 95%

confidence region and the solid line running through the studies represents the sROC curve.

In this graph the plotted results for each study fall a reasonable distance from the sROC curve suggesting heterogeneity in study results. Furthermore, only 2 of the possible 13 studies sit within the confidence region. Appendix VI details the data extracted and combined in meta-analysis.

Prevalence was calculated as previously described, using the formula:

$$\text{Prevalence} = \frac{\text{TP} + \text{FN}}{\text{TP} + \text{FN} + \text{FP} + \text{TN}}$$

$$\text{Prevalence} = \frac{302 + 125}{302 + 125 + 188 + 590}$$

In the present thesis, this equates to:

$$\text{Prevalence} = \frac{427}{1205}$$

$$\text{Prevalence} = 35\%$$

Therefore, the estimated prevalence of oropharyngeal aspiration in the present thesis is 35%.

The table below summarizes the data extracted from individual studies and combined in meta-analysis as detailed in Appendix VI. The table shows the combined results for true positives, false negatives, false positives and true negatives.

Table 4 A 2x2 classification table constructed using data extracted from the included studies.

	Reference standard positive (+) (VFSS)	Reference standard negative (-) (VFSS)	Total
Index test (CSA) (+)	289	195	484
Index test (CSA) (-)	138	583	721
Total	427	778	1205

In this thesis the overall sensitivity and specificity are calculated to be:

Overall sensitivity= 71%

Overall specificity= 76%

The positive predictive value (PPV) and negative predictive value (NPV) were also calculated using data in Table 4.

$$PPV = \frac{289}{484}$$

PPV= 60%

$$NPV = \frac{583}{721}$$

NPV= 81%

4.1.8 Sources of heterogeneity

This thesis explicitly identified and explored possible sources of heterogeneity, a process which has been found to be under reported in studies of DTA.⁸⁶ Given that visual examination of the sROC plot alone is an insufficient measure of heterogeneity of study results⁸⁶ each potential source of heterogeneity was identified and explored. Sources of heterogeneity relevant to studies of DTA include: design differences,

participants, variation in index test methods and study quality.^{13, 86} A further source of heterogeneity unique to analysis of diagnostic tests is variations in cut off values determined to indicate test positivity, leading to so-called ‘threshold effects’.⁸⁶ In studies of DTA, test results are presented in a binary manner, either test positive or negative, or disease present or absent.^{13, 86} However, in practice, a test result could be used to quantify the severity of a disease or condition, for example, mild/moderate/severe. In this instance the test effectively produces a continuous data set and it is the cut-off point along this data set that is used to define disease presence or absence by the person performing the test. Imaging tests, such as VFSS, can be reported on such a scale ranging from ‘definitely normal’ to ‘definitely abnormal’ with a range of severity categories between these two points. In this instance, thresholds can quickly become affected by inter-observer variation.⁸⁶ For example, one clinician may report a ‘moderate amount of aspiration’ for the same image a second clinician may report ‘trace amounts of aspiration’. In studies of DTA, the cut-off or threshold chosen to indicate a positive or negative test result may vary between studies of the same test⁸⁶, a finding that was supported in this thesis. In this thesis, the potential for threshold effects was tested by examining the sROC plot and comparing the cut off scores (used in each included study to determine a positive or negative test result) of different studies along the curve. The possibility of a threshold effect was tested by summarising the study results and index test method in a table to determine whether there was any systematic differences between the methods used for the index test. This data is presented and discussed in the discussion section of this review. See Table 6 for comparison of cut off scores used between studies.

4.1.9 Sub-group analysis

Sub-group analysis was not possible for the two potential groups identified in section 3.1.2 of the systematic review protocol as these sub groups were not available for comparison in the 13 included studies. Four additional potential sub-group analyses were identified *post-hoc*: the

health professional performing the CSA, results of aspiration for fluids and/or solids, time post onset of stroke diagnosis and the study quality. However, statistical sub-group analysis was not possible due to the relatively small number of included studies, as a minimum of five studies would have been needed per group.¹³ In the absence of formal statistical analysis, data from potential sub groups were analysed with the view to describing any apparent trends.

- Population of adults versus children: as described earlier, the one study⁵⁹ that included infants and children only, presented the lowest specificity for the clinical swallow assessment. This highlights the potential difficulty in recognising infants and children who are not aspirating. This population included children with neurological disease. No separate data is available at the moment for neurologically intact and otherwise normally-developing infants and children with dysphagia. This is a sub-group which requires further research.
- Health professional performing the clinical swallow assessment: the health professional performing the clinical swallow assessment varied between studies as did the clinician's level of experience working in the field, although in some studies this information was not explicitly stated. Clinicians included: medical staff^{92, 98}, dentists⁹⁸, speech pathologists^{59, 89-91, 93-97} and in two studies this information was not explicitly stated.^{61, 88} No overt trend in results was observed when comparing studies with a speech pathologist performing this assessment compared to other health clinician.
- Results for fluids and/or solids: as described in table of included studies (Appendix VII), some studies included fluids only in their results and some included fluids and solids. No overt trend identified in results for fluids only, compared with fluids and solids.
- Onset of stroke diagnosis: as described in the table of included studies Appendix VII the time since onset of stroke varied between papers. Five studies included patients with a stroke diagnosis of < 6

weeks and one paper included patients with a diagnosis > 6 weeks. No overt trend in results was observed when comparing studies with patients with a recent diagnosis of stroke (< 6 weeks) compared to > 6 weeks.

- Study quality: as described in table of included studies, there was variation in methodological quality between included studies. Higher quality papers (defined as QUADAS score 12 or greater) were compared with lower quality papers (defined as QUADAS score less than 12). There was no overt trend in results when these two groups were compared.

4.1.10 Summary of findings

This thesis has calculated the overall summary estimate of the diagnostic test accuracy of clinical swallow assessment for oropharyngeal aspiration as described below.

Patients/Population: mostly adults post CVA, only one study presented data for infants and children only (n=59) and one study included children under 15 years (n=10). The total number of participants from all 13 included studies was 1205.

Setting: mainly hospitals and rehabilitation centres.

Index Test: clinical swallow assessment, performed by a range of health professionals including physicians, dentists, occupational therapists and speech-language pathologists.

Reference standard: video fluoroscopic swallow study, performed predominantly by radiologists and speech-language pathologists.

Studies: predominantly cross-sectional prospective studies.

Table 5 Summary of study findings

Mean sensitivity and specificity	Positive and negative predictive value	Number of patients (number of studies)	Prevalence	What do these results mean?
Sensitivity: 71% Specificity: 76%	PPV: 60% NPV: 81%	1205 (13)	35%	35/100 patients are actually aspirating. Of these, 25 are correctly identified by clinical swallow assessment (CSA) and 10 are missed (false negatives). 49 out of 65 patients are correctly identified by CSA as not aspirating, the remaining 16 are incorrectly identified as aspirating (false positive). Using the PPV and NPV results, this means of 60% of people who test positive for aspiration by CSA will be aspirating and 81% who test negative on CSA will not be aspirating

Figure 6 demonstrates the clinical significance of receiving a positive or negative test result given the calculated PPV and NPV values. These values are based on the study population presented in this thesis which is predominantly adult, acute post stroke patients. The values are based on a calculated prevalence of 35%.

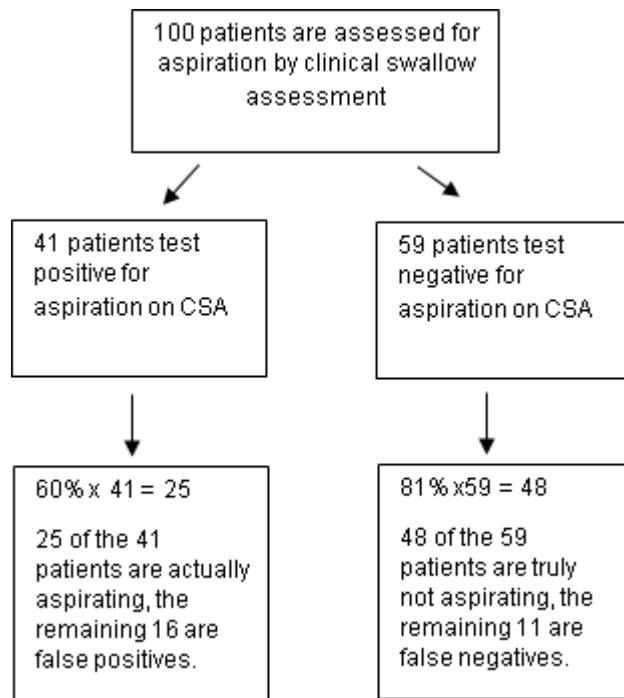


Figure 6 This flowchart demonstrates the likelihood of a test result being accurate differs for a positive or negative test result

5 Discussion

5.1 Discussion

This thesis has compared the diagnostic test accuracy of clinical swallow assessment (CSA) with video fluoroscopic swallow study (VFSS). VFSS is considered the 'Gold Standard' for assessment and diagnosis of oropharyngeal aspiration ^{22, 56} and used frequently as the reference standard in tests of diagnostic test accuracy. ^{20, 42, 58, 59} However, this technique is not always available and therefore the CSA is considered an alternative.

5.1.1 Summary of main results

This thesis included 13 homogeneous studies that were subsequently combined in meta-analysis. Included studies explicitly reported the use of CSA compared with VFSS for assessment and diagnosis of oropharyngeal aspiration. The majority of studies presented with reasonably high methodological quality scores.

5.1.2 Findings of meta-analysis

This thesis has combined the data extracted from the 13 included studies in meta-analysis to calculate a summary estimate of the test accuracy of clinical swallow assessment. Figure 4 presents this data as a forest plot and demonstrates that sensitivity in particular, is heterogeneous ranging from 21% - 93%, whereas specificity ranges from 46% - 93%. Figure 5 presents data from all included studies as an sROC plot, and calculates the overall sensitivity and specificity of CSA as 71% and 76% respectively. Using this data the positive and negative predictive values were also calculated and discussed below. The appearance of this data suggests heterogeneity in study results, as many of the data points for each included study do not fall on the sROC curve and the relatively wide

confidence intervals of the summary estimate (especially for sensitivity) ^{13, 86}. In addition, the shape of the curve is strongly suggestive of a threshold effect, with sensitivity and specificity varying as different studies use different cut-off points or interpret the subjective CSA differently. The issue of heterogeneity is discussed further in the following section.

5.1.3 Population and settings

One of the aims of this thesis was to review the diagnostic test accuracy of CSA compared with VFSS for infants and children, however, the most common patient group presented in the literature was adults presenting with first-ever acute stroke. One study included some children less than 15 years of age ⁹¹ and only one study exclusively presented results for infants and children. ⁵⁹ The cause for dysphagia was a neurological aetiology in nearly all of the patients in the included studies. A small number of infants and/or children in the DeMatteo study ⁵⁹ may have been neurologically intact as diagnoses included cardiac conditions and failure to thrive, although numbers for subgroups were not provided. The results from this thesis therefore reflect a very specific patient group – adults post-acute stroke which may not be generalised to other patient groups such as infants, children or neurologically intact infants, children and adults.

5.1.4 Reference test methods

Included studies provided varying amounts of detail regarding the methods used during VFSS. Nine studies were considered to provide sufficient detail to replicate the study, four studies did not provide sufficient details to permit replication. ^{89, 94, 98, 99} The studies that did include sufficient information stated the specifics of the study such as: patient position/seating, contrast used (for example, liquid Barium compared with Omnipaque), amount of food and fluid offered (eg 5mL, half teaspoon), the order of presentation of varying boluses and the health professionals who interpreted the study results.

Only one study ⁹³ specified the recipe used to create the various consistencies used during the VFSS. For example 'thin liquid barium' can be created using Barium with water, however the more water that is used, the 'thinner' the substance which more accurately reflects the water given during CSA.

All studies stated the equipment used and that a lateral view of the patient was obtained, some studies also included an antero-lateral view. A lateral view is most frequently used for assessment of oropharyngeal aspiration. Each study used a different recording system which may impact on the clarity of images and the functions of the recording systems to play back the films in slow motion and assess for evidence of smaller amounts (micro or trace) of aspiration.

5.1.5 Index test methods

Included studies provided varying amounts of detail regarding the methods used during the CSA and there is no standard method used by clinicians during CSA. Practice varies within and between centres. Ten studies were considered to provide sufficient detail to replicate the study, three studies did not provide sufficient details to permit replication. ^{89, 90, 99} The studies that did include sufficient information stated the specifics of the study such as: patient position/seating, sequence of the assessment for example, oral examination followed by oral trials, the types and quantities of oral trials and the rating used to include or exclude the diagnosis of aspiration.

Only two ^{61, 93} of the included studies used the exact amount and consistency of oral trials during the CSA that was used in the VFSS. For example, in the McCullough 2001 study ⁶¹ participants were offered two swallows of each consistency – thin, thick liquid, puree and solid in 5 ml amounts and ¼ of a cookie. These exact bolus measurements were then offered during the VFSS. In the remaining studies, the trials offered during CSA were either different to those offered in VFSS or there was insufficient information provided.

5.1.6 Effect of fatigue and aspiration

None of the included studies discussed or accounted for the timing of aspiration during the CSA or VFSS, that is, if the aspiration occurred at the beginning, middle or end of the study. Aspiration in infants often does not occur in the first few swallows during a VFSS but rather, after the infant has had multiple fluid swallows.⁵² Studies reported the presence or absence of aspiration but did not report the time length of the study or the number of x-ray films recorded and the timing of the aspiration during the study. During VFSS, one strategy to factor in the effect of fatigue during swallowing is to allow the infant to continue to feed and only intermittently screen using fluoroscopy. This can allow for the possible impact of fatigue during feeding without exposing the infant to unnecessary radiation and fluoroscopy recording. If the infant's swallow is only assessed at the beginning of their feeding the diagnosis of aspiration may be missed as the aspiration event may occur later during the feeding process.

5.1.7 Heterogeneity of study results

The following section identifies and discusses the possible sources of heterogeneity between study results, as indicated by the distance of a study (as represented by its oval marker) from the sROC curve (Figure 5).

Design differences

Study design differences can include methods used to recruit participants, the study setting, prospective or retrospective study and whether or not all participants received the same reference standard. When all participants receive the same reference standard, this is considered the strongest study design.¹⁰²

There were differences between studies in the methods used to recruit patients, the main difference was that some studies recruited patients based only on the diagnosis of stroke, while other studies recruited patient based on symptoms of dysphagia. Specifically, five of the included studies

included participants based on their diagnosis of stroke, without mention of any swallowing or dysphagic symptoms.^{61, 93-96} In six of the included studies, the patients were referred specifically for swallow assessment and presenting with one or more features of dysphagia.^{59, 90-92, 97, 98} In two of the studies this information was not reported.^{89, 99} No obvious trend was seen on the sROC plot for studies that recruited participants on the basis of stroke diagnosis compared with symptoms of dysphagia.

Study setting may impact on study results, for example if some studies are set in a primary health facility compared with a secondary or tertiary facility. The most common setting of the included studies was an acute care hospital – six of the included studies were set in a hospital. The potential implications of this are that the participants may present with acute dysphagic symptoms as opposed to chronic or more stable symptoms which could influence the CSA results. For example an acute stroke patient is likely to present with severe dysphagic symptoms compared to a longer-term stroke patient.¹⁰⁴ Three of the five studies set in an acute hospital setting^{92, 95, 98} presented with reasonably high paired sensitivity and specificity and can be seen on the sROC plot as the three studies closest to the upper left corner of the plot.

Another potential difference between study designs is whether or not all participants received the same reference standard. This was not identified as a factor in this thesis as all included studies used the same reference standard – VFSS.

Participants

Heterogeneity was identified for study participants between the included studies and none of the included studies stated the severity of symptoms of dysphagia. For example, a participant with severe symptoms of dysphagia such as coughing, choking and dehydration presents with a different swallowing profile compared to a participant with mild symptoms of dysphagia or no symptoms of dysphagia. The table of included studies provides details of study participants (Appendix VII), however in summary the main differences included: time post onset of dysphagia, time post

onset of stroke diagnosis, neurological cause of dysphagia and the age range of participants which ranged from 18 – 96 years of age in the adult studies.

Variation in index test methods

Differences in the methods used for the index test can lead to heterogeneity in study results. Index test methods did vary slightly between the included studies. One particular consideration in studies of DTA is whether or not a screening or diagnostic tool is used.⁸⁶ Two of the included studies^{92, 97} used a screening tool for the CSA in comparison to the remaining 11 studies which used a complete CSA tool. Despite the remaining studies using a comprehensive CSA, differences in the index test were identified such as: the type and amount of food/fluid used to test the swallow, the check list used to guide the assessment and the scoring system used to identify the presence or absence of aspiration as summarised in the table of included studies (Appendix VII).

Study quality

A potential source of heterogeneity in studies of DTA is study quality and risk of bias.⁸⁶ As discussed previously, the overall quality of included studies was reasonably high. However, as described in Figure 3, potential bias was identified in certain areas of study quality. For example, in eight of the included studies the time delay between the index test and reference test was either not reported or not sufficient. The implications of this on study results are that the severity of the dysphagia or aspiration may have changed – improved or worsened – during the time between the index test and reference test was performed. One study identified with a reasonably poor overall quality score of 5/14 using the QUADAS checklist did not report critical information such as: selection criteria, whether or not the index test results were blinded to the reference test and did not explicitly state which patient sample received the reference standard. This was identified as a heterogeneous study as it fell a reasonable distance from the sROC curve. Poor quality studies may introduce heterogeneity in study results due to bias.

Cut-off values and threshold effects

A potential source of heterogeneity in study design during analysis of diagnostic tests is variations in cut off values determined to indicate test positivity, as the threshold chosen may vary between studies of the same test.¹³ Implicit variation is expected to exist between included studies due to the nature of the index test and the reference test, which both relied on clinical interpretation, and expected variation in the worldwide use of this assessment tool within and between hospitals and health care facilities. Figure 5 shows that the shape of the curve on the sROC plot is consistent with a threshold effect.^{13, 86} Included studies used a binary outcome to score the index test – presence or absence of aspiration. This classification of ‘present’ or ‘absent’ depends on whether the clinician performing the test measures a given trait as above or below a defined cut off or threshold value. The threshold chosen may vary between studies of the same test.⁸⁶ One study furthest from the summary estimate of the sROC (Figure 5) is Splaingard et al⁹¹. The Splaingard et al study was one of two studies set in a rehabilitation centre rather than a hospital. The population did include CVA patients, however, it also included participants with brain injury and chronic neuromuscular disease. The population included in Splaingard et al study– closed head injury and neurodegenerative pathologies, is a patient group with one of the highest rates of silent aspiration¹⁰⁵ which could explain the relatively low sensitivity of 20%.

Table 6 presents the method of scoring the clinical swallow assessment as well as chosen thresholds. As this table illustrates, the threshold used to score aspiration present or absent relied on ‘overall clinical judgment’ of the swallow quality in approximately half of the included studies. Six of the included studies used a pre-determined criteria to rate the patient as aspirating or not aspirating. These inter-study differences contribute to the overall heterogeneity of the included studies and the resulting appearance of the sROC plot.

Table 6 This table presents the between study differences and similarities used to rate the presence or absence of aspiration during clinical swallow assessment and the cut-off values used to determine the score.

Study	Scoring of clinical swallow assessment	Cut-off for scoring
Barbiera 2006	Three possible ratings, only two relating to aspiration either: dysphagia and aspirating; dysphagic and not aspirating; not dysphagic.	Based on overall judgment of clinical swallow assessment.
Baylow 2009	Dichotomous scoring system, patient rated by clinician with each consistency as either: safe or unsafe. Did not explicitly state the patient was assessed as aspirating however this was inferred from the use of the terms 'safe' (not aspirating) and 'unsafe' (aspirating).	Based on 'preset criteria' – criteria not clearly described.
DeMatteo 2005	Clinician's level of confidence about suspicions of aspiration presented as a percentage, three possible ratings: 100% -certain aspirating; 0% - certain aspiration not present; 50% - uncertainty regarding this decision.	Based on overall judgment of clinical swallow assessment.
DePippo 1992	Dichotomous scoring system – scored as 'normal' or 'abnormal'.	Scored as abnormal if coughing during or one minute after water swallow test, or presence of post-swallow wet-hoarse voice quality.
Hammond 2009	Dichotomous scoring system – classified as aspiration present or absent.	Scored as aspirating if any of the three assessment features of the CSA were judged as abnormal.
Linden 2009	Dichotomous scoring system – categorised as either: exhibiting subglottic penetration (aspiration) or no evidence subglottic penetration.	Not described
Mann 2000	Four possible ratings. Presence of aspiration either: unlikely; possible; probable; definite.	Unlikely – no detected abnormality. Possible – at least one component of swallow affected. Probable – several components of swallow affected. Definite – several items of swallow assessment affected (>5) and considerable risk airway compromise.
McCullough 2001	Dichotomous scoring system – classified as aspiration present or absent.	Based on overall estimate of the clinical swallow assessment.
McCullough 2005	Dichotomous scoring system – classified as aspiration present or absent.	Based on overall estimate of the clinical swallow assessment.
Smithard 1998	Dichotomous scoring system, patient rated by clinician with each consistency as either: safe or unsafe. Did not explicitly state the patient was assessed as aspirating however this was inferred from the use of the terms	Based on overall estimate of the clinical swallow assessment.

	'safe' (not aspirating) and 'unsafe' (aspirating).	
Splaingard 1988	Five point scale, only two relating to presence of aspiration, the other three related to overall swallow function: 1 – within normal limits; 2 mild – slight delay swallow reflex; 3 moderate – delayed swallow, post-swallow residue; 4 severe – trace aspiration; 5 – profound - >10% aspiration.	Based on overall estimate of the clinical swallow assessment.
Tohara 2003	Five point scale. Test potentially terminated at a score of 3. If progressed beyond this to '4' or '5' patient needed to be able to 'pass' the next step. Test repeated twice, final score = lowest score on any trial.	Score >3 was defined as 'abnormal'. This cut off value was tested at 4 points by the authors and this score had the highest concordance ratio.
Zhou 2011	Two part assessment. Part one used a scoring system constructed using discriminant analysis. Six different items of the CSA with an assigned score. If patient received a score of 14-28 (out of possible 42) – recorded as 'inconclusive'. This group then went on to receive 3-oz water swallow test. Water swallow test used dichotomous scoring of aspiration present or absent.	Scored as aspirating if: the task not completed, or if coughing, choking or wet-hoarse voice present. Voice measured during, within and one minute after the end of the test.

5.1.8 The diagnostic test accuracy of elements of the CSA

A number of studies (5) reported the DTA of elements of the CSA separately. Table 7 provides a comparison of the salient features of the CSA as reported in the included studies. In instances where 'overall judgment of aspiration' was reported as a clinical feature, this has not been included as this is not a specific, measurable feature of the CSA that may assist clinicians when implementing CSA as a diagnostic tool for aspiration.

Table 7 Diagnostic accuracy of various features of CSA as presented in 5 of the 13 included studies. A 'tick' symbol represents this information was present in the study.

Study	Element of CSA	Sensitivity (percentage)	Specificity (percentage)	Significant feature (p < 0.05)
Baylow et al. ⁹⁰	Cough/throat clear	67	33	Not measured
Linden et al. ⁸⁹	Wet spontaneous cough	Not measured	Not measured	✓
McCullough et al. ⁶¹	Cough during swallowing	68	81	Not measured
McCullough et al. ⁹³	Spontaneous cough	44	82	Not measured
McCullough et al. ⁹³	Spontaneous throat clear	54	69	Not measured
Baylow et al. ⁹⁰	'Gurgly' voice	67	50	Not measured
Linden et al. ⁸⁹	Wet phonation	Not measured	Not measured	✓
McCullough et al. ⁹³	Wet voice	63	64	Not measured
McCullough et al. ⁹³	Cough/throat clear/wet voice (combined)	81	47	Not measured
Smithard et al. ⁹⁴	*Weak voluntary cough *Any impairment level of consciousness *presence of one or both of these	75	72	Not measured
McCullough et al. ⁹³	Delayed oral transit	56	71	Not measured
Linden et al. ⁸⁹	reclined position	Not measured	Not measured	✓
Linden et al. ⁸⁹	dysphonia or aphonia	Not measured	Not measured	✓
Linden et al. ⁸⁹	abnormal gag	Not measured	Not measured	✓
Linden et al. ⁸⁹	Impaired swallowing on secretions	Not measured	Not measured	✓
Linden et al. ⁸⁹	Altered phonation	Not measured	Not measured	✓

As described in the Introduction section of this thesis, 'cough' is a frequently used clinical sign to assist with the diagnosis of aspiration. The findings of this thesis demonstrate the variable diagnostic accuracy of cough/throat clear for diagnosis of oropharyngeal aspiration. Table 7 shows the sensitivity of cough/throat clear ranged from 44%-68%, to summarise, this sensitivity is quite low.

The feature of CSA with the highest sensitivity (closest to 100%) was cough/throat clear/wet voice combined – sensitivity 81%. This finding is consistent with a recent retrospective study which showed the combination of cough, wet voice and wet breathing was significantly associated with oropharyngeal aspiration of thin fluids in children.⁵⁴ However, this cluster of clinical signs has a relatively low specificity – 47%. This finding has been demonstrated by others in the literature – an absence of cough does not necessarily indicate an absence of aspiration.¹⁰⁶ Clinicians may perceive a false positive test result to be ‘safer’ than a false negative test result due to the risk of compromised lung health if patient receives a false negative test result.⁹⁸ However as will be discussed in the implications for practice section of this paper, a false positive test result can have a negative impact on patient management, particularly if the prescribed, modified diet poses a risk of silent aspiration.

5.1.9 Prevalence and positive and negative predictive values

Positive and negative predictive values are affected by the prevalence of the condition in the population. For example, Figure 6 describes the likelihood of a test result being accurate based on a prevalence of 35% (the prevalence in the present study). If the prevalence increases to 60%, the PPV increases to 81% (43/53) and NPV decreases to 64% (30/47). This means in a population where aspiration is more prevalent, the CSA test is better at correctly identifying aspiration in those who are actually aspirating. In contrast, the ability of the CSA to correctly identify those who are not aspirating decreases with an increase in prevalence. The tables below demonstrate the relationship between prevalence and PPV and NPV.

Table 8 Comparison of PPV and NPV using varying prevalence of aspiration.

Prevalence 10%	Prevalence 60%
PPV: 24%	PPV: 81%
NPV: 96%	NPV: 64%
This means that only 24% of patients that test positive for aspiration on CSA are actually aspirating. The remaining patients are false positives. 96% of patients that test negative for aspiration on CSA are truly not aspirating. The remaining are false negatives.	This means that 81% of patients that test positive for aspiration on CSA are actually aspirating. The remaining patients are false positives. 64% of patients that test negative for aspiration on CSA are truly not aspirating. The remaining are false negatives.

Table 8 shows that overall the PPV increases and the NPV decreases as prevalence increases and the PPV decreases and the NPV increases as prevalence decreases. This means that clinicians should have greater confidence in their diagnosis of excluding aspiration in a lower prevalence population but less confidence in accurately diagnosing the presence of aspiration in this same lower prevalence population. Therefore, understanding of the prevalence of the condition is very important to understanding a diagnostic test result. This thesis provides a calculated prevalence of aspiration which has not been reported previously. The prevalence of aspiration was calculated to be 35%.

Below are the calculations used to obtain the figures in Table 8.

Table 9 A 2x2 classification used to determine PPV and NPV for prevalence of 60%

Index test	Reference standard positive (+)	Reference standard negative (-)	Total
(+)	43	10	53
(-)	17	30	47
Total	60	40	100

The calculations of PPV and NPV based on a sensitivity of 71%, a specificity of 76% and a prevalence of 60% are:

Overall sensitivity = 71%. $0.71 \times 60 = 43$

Overall specificity = 76%. $0.76 \times 40 = 30$.

The positive predictive value (PPV) and negative predictive value (NPV) were conducted using Table 9

PPV = $43/53 = 81\%$

NPV = $30/47 = 64\%$

Table 10 A 2x2 classification used to determine PPV and NPV for prevalence of 10%.

Index test	Reference standard positive (+)	Reference standard negative (-)	Total
(+)	7	22	29
(-)	3	68	71
Total	10	90	100

The calculations of PPV and NPV based on a sensitivity of 71%, a specificity of 76% and a prevalence of 10% are:

Overall sensitivity = 71%. $0.71 \times 10 = 7$

Overall specificity = 76%. $0.76 \times 90 = 68$

The positive predictive value (PPV) and negative predictive value (NPV) were conducted using Table 10

PPV = $7/29 = 24\%$

NPV = $68/71 = 96\%$

6 Conclusions

6.1.1 Implications for practice

The results from this thesis are relevant to a reasonably specific population – adults post stroke and can not necessarily be generalised to other patient groups. This thesis demonstrates that in a population with a relatively high prevalence of aspiration (35%) a clinician's assessment regarding absence of aspiration using a CSA is more likely to be accurate than their assessment of the presence of aspiration. There are implications for a patient care of receiving a false positive or false negative test result, which must be weighed up when deciding whether to administer a test and how to interpret its findings.

Implications of a false positive test result

A false positive test result for an adult may result in a range of poor outcomes that cause patient and/or family distress and may induce unnecessary care and costs. A false positive test result may result in withholding oral medications. Oral medications such as tablets and capsules are considered a choking risk in the dysphagic population and are often with-held if there is concern regarding swallowing safety. Certain medications are not able to be crushed or given in liquid form which means there may be a period of time patients are not receiving necessary medications.

A false positive test result will likely result in ceasing the patient's usual fluids and food and commencing a strict modified diet and/or fluids. Modification of fluids usually involves prescribing thickened fluids and a modified diet usually involves restricting the oral diet to solids that require less chewing. Unnecessary diet and fluid modification may result in weight loss and/or dehydration. Diet modification may lead to reduced oral feeding skills due to the patient no longer chewing solids and needing to manipulate a bolus prior to swallowing and reduced quality of life by restricting the enjoyment of 'normal' food.

If an adult patient receives a false positive CSA test result and is considered to be aspirating all food and liquid consistencies, they may be considered for enteral feeding and recommended nil by mouth, which again would constitute unnecessary care and cost. In addition the patient may experience distress and dissatisfaction with this recommendation.

A false positive test result for an infant or child may also result in modified oral diet and fluids. In a breast feeding infant, the recommendation may be to cease breast feeding as this thin liquid is unable to be thickened when fed from the breast. The infant's formula is likely to be thickened which can be problematic if the infant's sucking skills do not allow the thickened fluid to draw through the bottle teat. Furthermore, unnecessary modification of the infant or child's oral diet may delay development of oral chewing and swallowing skills. Aspiration of thicker consistencies has been shown to increase the risk of developing pneumonia when compared with aspiration of thin liquids.¹⁰⁷ This may be due in part to the fact that children with moderate or severe dysphagia are placed on thicker fluids. Nonetheless it is a concern for the children placed on thicker fluids without the sufficient diagnostic information to confirm the presence or absence of aspiration.

The infant or child's initial presenting symptoms will not be investigated if the diagnosis of aspiration is made and considered the cause of the original symptoms. This may delay appropriate diagnosis and treatment for the infant or child.

Implications of a false negative test result

A false negative test result for an adult will result in the patient continuing to aspirate which may lead to compromised lung health or pneumonia. Aspiration pneumonia can be fatal in certain patient populations, such as medically complex patients with associated respiratory conditions.¹⁰⁸ Furthermore the patient may be at risk of choking with solid food which is also potentially fatal. In addition to the direct impact on patient management, there are broader aspects of patient care that may be affected and induce a cost to the health care facility. For example, further

invasive tests and investigations may be arranged to assist with diagnosis of the initial presenting symptoms. A delay in medical and/or multi-disciplinary treatment and commencement of therapy will also occur.

A false negative test result for an infant or child places them at risk of chronic pulmonary aspiration – the repeated entry of material into the subglottic airway potentially leading to chronic or recurrent respiratory symptoms.^{21, 54} Aspiration left untreated can also result in progressive lung disease, bronchiectasis and in severe cases respiratory failure²² In addition to these potential respiratory sequelae, the infant or child may experience malnutrition and stressful feeding interactions with care givers⁴² due to the discomfort and distress experienced during feeding. Furthermore the patient may not be referred to the necessary paediatric speciality services and multi-disciplinary feeding/swallowing programs.⁴²

Children who are repeatedly fed during a sub-optimal feeding state may go on to develop an oral aversion, experience failure to thrive and contribute to increased stress between the care-giver and the child/patient. Infants and children fed while they are unknowingly aspirating during feeds often present with poor feeding skills including: refusal, distress and pulling away from the bottle/breast. Due to the increased time required by parents to spend feeding the infant to achieve sufficient growth, nutrition and hydration, it is probable that there would be less time available for other important developmental activities such as play and development of early communication and motor skills.

Implementation of CSA

Table 6 presents the between study differences and similarities used to rate the presence or absence of aspiration during clinical swallow assessment and the cut-off values used to determine the score. In the majority of studies, clinicians used their 'overall clinical judgment' to determine the presence or absence of aspiration. During CSA, clinical reasoning and expertise underpins the clinician's final determination regarding whether or not a patient is aspirating. In settings where VFSS is not available it is essential that clinicians performing the CSA have access

to the necessary training and support to perform the CSA in order to confidently determine whether or not the patient is aspirating. In settings where VFSS is unavailable, clinicians could consider using a numbered rating to determine the presence or absence of aspiration rather than an 'overall clinical judgment'. Use of a numbered rating for each component of the CSA then calculating a final score would assist in determining which features of the CSA were indicators of the presence of aspiration, if there is certainty about the presence of aspiration. Use of a numbered rating could also reduce the subjectivity of CSA and enable clinicians to potentially 'adjust' their scoring system when comparing results of the CSA with results of the VFSS. The CSA was different for all included studies. All studies used different check lists to guide the assessment and some studies did not provide or site the checklist used. All included studies offered different oral trials, in a different order. Centres using CSA for assessment of oropharyngeal aspiration should consider standardising the assessment tool to reduce the potential impact of clinical bias.

6.1.2 Concluding Summary

In centres where VFSS is unavailable this thesis provides data to assist in the interpretation of CSA results. To summarise, data was extracted from 13 included studies and combined in meta-analysis. The overall sensitivity of CSA compared with VFSS was 71% and the mean specificity was 76%. Positive and negative predictive values were calculated as 60% and 81% respectively and the prevalence of aspiration was calculated as 35%. Using this data and prevalence information, this means that 60% of patients identified as aspirating on CSA truly are aspirating and 81% who test negative for aspiration on CSA are not aspirating. Clinicians should use this diagnostic information to guide their decision-making during CSA in situations where VFSS is not available or not practical.

It is recommended that hospitals and centres assessing patients for oropharyngeal aspiration consider accessing a facility where VFSS is available. In particular when assessing for oropharyngeal aspiration in a

lower prevalence population. VFSS is a fundamental component of the accurate diagnosis of aspiration, and prescribing effective treatment approaches for the management and treatment of oropharyngeal aspiration.

6.1.3 Implications for research

This thesis highlights the lack of data available for diagnostic test accuracy of clinical swallow assessment in the diagnosis of aspiration for infants and children. This thesis demonstrates that only one paper is available for this population. There is no data available for neurologically intact and normally developing children, which is another area requiring research.

6.1.4 Limitations of the review

As described in the 'results' section of this thesis some data was extracted by the author to populate the 2x2 tables in instances where the information was not provided in the original paper. Although measures were taken to limit bias such as reviewing this data extraction with an independent reviewer, this does present a possible source of bias. Furthermore, studies were limited to English language published up until April 2012.

7 References

1. Jackson N. Systematic reviews of health promotion and public health interventions handbook. Available from: ph.cochrane.org/sites/ph.cochrane.org/files/uploads/HPPH_systematic_review_handbook.pdf.
2. National Health and Medical Research Council. A guide to the development, implementation and evaluation of clinical practice guidelines. Canberra: Australian Government Publishing Service; 1998 [cited 2013 7/11/2013].
3. Pearson A, Wiechula R, Court A, Lockwood C. The JBI model of evidence-based healthcare. *International Journal of Evidence Based Healthcare*. 2005;3(8):207-15. Epub 2005/09/01.
4. Hemingway P, Brereton N. What is a systematic review? 2009 [[2/11/2013]]; 2nd:[Available from: <http://www.medicine.ox.ac.uk/bandolier/painres/download/whatis/syst-review.pdf>].
5. Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC, Ioannidis JPA, et al. The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: Explanation and elaboration. *PLoS Medicine*. 2009;6(7).
6. The Cochrane Library. The Cochrane Library publishes the most thorough survey of MMR vaccination data which strongly supports its use. [7/11/2013]; Available from: <http://download.pdf-world.net/The-Cochrane-Library-publishes-the-most-thorough-survey-of-MMR-download-w8687.html>.
7. Wakefield AJ, Murch SH, Anthony A, Linnell J, Casson DM, Malik M, et al. Ileal-lymphoid-nodular hyperplasia, non-specific colitis, and pervasive developmental disorder in children. *Lancet*. 1998;351(9103):637-41. Epub 1998/03/21.
8. Murch SH, Anthony A, Casson DH, Malik M, Berelowitz M, Dhillon AP, et al. Retraction of an interpretation. *Lancet*. 2004;363(9411):750.
9. Demicheli V, Jefferson T, Rivetti A, Price D. Vaccines for measles, mumps and rubella in children. *Cochrane database of systematic reviews* (Online). 2005(4).
10. The Cochrane Collaboration. About cochrane systematic reviews and protocols. John Wiley & Sons; 2013 [7/11/2013].
11. Mulrow CD. The Medical Review Article: State of the Science. *Annals of Internal Medicine*. 1987;106(3):485 - 8.
12. Leeflang MMG, Deeks JJ, Gatsonis C, PMM B. Systematic reviews of diagnostic test accuracy. *Annals of Internal Medicine*. 2008;149 (12):889-97

-
13. White S, Schultz T, Enuameh YAK. Conducting a systematic review of diagnostic test accuracy evidence. In: Professor Pearson A, editor. Synthesizing evidence of diagnostic accuracy. Philadelphia: Lippincott Williams and Wilkins; 2011. p. 49-73.
 14. Dubourg J, Berhouma M, Cotton M, Messerer M. Meta-analysis of diagnostic test accuracy in neurosurgical practice. *Neurosurgical Focus*. 2012;33(1):E5.
 15. Bossuyt PM, Irwig L, Craig J, Glasziou P. Comparative accuracy: Assessing new tests against existing diagnostic pathways. *British Medical Journal*. 2006;332(7549):1089-92.
 16. Gordis L. *Epidemiology*. Third ed: Elsevier Inc.; 2004.
 17. Jones CM, Athanasiou T. Diagnostic accuracy meta-analysis: Review of an important tool in radiological research and decision making. *British Journal of Radiology*. 2009;82(978):441-6.
 18. Weinstein S, Obuchowski NA, Lieber ML. Clinical evaluation of diagnostic tests. *American Journal of Roentgenology*. 2005;184(1):14-9.
 19. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996;11(2):93-8.
 20. Weir K, McMahan S, Barry L, Ware R, Masters IB, Chang AB. Oropharyngeal aspiration and pneumonia in children. *Pediatric Pulmonology*. 2007;42(11):1024-31.
 21. Boesch R, Daines C, Willging JP, Kaul A, Cohen AP, Wood RE, et al. Advances in the diagnosis and management of chronic pulmonary aspiration in children. *European Respiratory Journal*. 2006;28(4):847-61.
 22. de Benedictis FM, Carnielli VP, de Benedictis D. Aspiration lung disease. *Pediatric Clinics of North America*. 2009;56(1):173-90.
 23. Dodds WJ. The physiology of swallowing. *Dysphagia*. 1989;3(4):171-8.
 24. Logemann JA. Swallowing physiology and pathophysiology. *Otolaryngologic Clinics of North America*. 1988;21(4):613-23.
 25. Arvedson JA, Brodsky L. *Pediatric swallowing and feeding assessment and management*. Second ed. USA: Delmar Cengage Learning; 2002.
 26. Logemann JA. *Evaluation and Treatment of Swallowing Disorders*. 2nd ed. USA: Pro-Ed Incorporated; 1998.
 27. Wilson JA, Pryde A, White A, Maher L, Maran AGD. Swallowing performance in patients with vocal fold motion impairment. *Dysphagia*. 1995;10(3):149-54.
 28. McConnel FMS. Analysis of pressure generation and bolus transit during pharyngeal swallowing. *Laryngoscope*. 1988;98(1):71-8.
 29. Marik PE. Aspiration pneumonitis and aspiration pneumonia. *New England Journal of Medicine*. 2001;344(9):665-71.

-
30. Croghan JE, Burke EM, Caplan S, Denman S. Pilot study of 12-month outcomes of nursing home patients with aspiration on videofluoroscopy. *Dysphagia*. 1994;9(3):141-6.
 31. Terpenning MS, Taylor GW, Lopatin DE, Kerr CK, Liza Dominguez B, Loesche WJ. Aspiration pneumonia: Dental and oral risk factors in an older veteran population. *Journal of the American Geriatrics Society*. 2001;49(5):557-63.
 32. Logemann JA. Treatment for aspiration related to dysphagia: An overview. *Dysphagia*. 1986;1(1):34-8.
 33. Altman KW, Richards A, Goldberg L, Frucht S, McCabe DJ. Dysphagia in Stroke, Neurodegenerative Disease, and Advanced Dementia. *Otolaryngologic Clinics of North America*. 2013;Article in press.
 34. Tutor JD, Gosa MM. Dysphagia and aspiration in children. *Pediatric Pulmonology*. 2012;47(4):321-37.
 35. Smith J, Wolkove N, Colacone A, Kreisman H. Coordination of eating, drinking and breathing in adults. *CHEST*. 1989;96(3):578-82.
 36. Barer DH. The natural history and functional consequences of dysphagia after hemispheric stroke. *Journal of Neurology Neurosurgery and Psychiatry*. 1989;52(2):236-41.
 37. Rofes L, Vilardell N, Clavé P. Post-stroke dysphagia: Progress at last. *Neurogastroenterology and Motility*. 2013;25(4):278-82.
 38. Martino R, Terrault N, Ezerzer F, Mikulis D, Diamant NE. Dysphagia in a patient with lateral medullary syndrome: Insight into the central control of swallowing. *Gastroenterology*. 2001;121(2):420-6.
 39. American-Speech-Language-Hearing Association. Swallowing disorders (dysphagia) in adults. USA2012 [cited 2012 20/05]; Available from: www.asha.org/public/speech/swallowing/SwallowingAdults.htm.
 40. Okubo PCMI, Fabio SRC, Domenis DR, Takayanagui OM. Using the national institute of health scale to predict dysphagia in the ischemic stroke. *Cerebrovascular Diseases*. 2012 33(6):501-7.
 41. Aviv JE, Martin JH, Jones ME, Wee TA, Diamond B, Keen MS, et al. Age-related changes in pharyngeal and supraglottic sensation. *The Annals of otology, rhinology, and laryngology*. 1994;103(10):749-52.
 42. Lefton-Greif MA, Carroll JL, Loughlin GM. Long-term follow-up of oropharyngeal dysphagia in children without apparent risk factors. *Pediatric Pulmonology*. 2006;41(11):1040-8.
 43. Prasse JE, Kikano GE. An overview of pediatric dysphagia. *Clinical Pediatrics*. 2009;48(3):247-51.
 44. Martino R, Foley N, Bhogal S, Diamant N, Speechley M, Teasell R. Dysphagia after stroke: incidence, diagnosis, and pulmonary complications. *Stroke*. 2005;36(12):2756-63.
 45. Arvedson JC. Assessment of pediatric dysphagia and feeding disorders: Clinical and instrumental approaches. *Developmental Disabilities Research Reviews*. 2008;14(2):118-27.

-
46. Brackett K, Arvedson JC, Manno CJ. Pediatric Feeding and Swallowing Disorders: General Assessment and Intervention. Perspectives on Swallowing and Swallowing Disorders (dysphagia). 2006;15(3):10-5.
 47. Reilly S, Skuse D, Poblete X. Prevalence of feeding problems and oral motor dysfunction in children with cerebral palsy: A community survey. The Journal of Pediatrics. 1996;129(6):877-82.
 48. Forsyth BWC, Leventhal JM, McCarthy PL. Mothers' perceptions of problems of feeding and crying behaviors. American Journal of Diseases of Children. 1985;139(3):269-72.
 49. Beautrais AL, Fergusson DM, Shannon FT. Family life events and behavioral problems in preschool-aged children. Pediatrics. 1982;70(5):774-9.
 50. Manikam R, Perman JA. Pediatric feeding disorders. Journal of Clinical Gastroenterology. 2000;30(1):34-46.
 51. Kim JS, Han ZA, Song DH, Oh HM, Chung ME. Characteristics of dysphagia in children with cerebral palsy, related to gross motor function. American Journal of Physical Medicine and Rehabilitation. 2013;92(10):912-9.
 52. Newman LA, Keckley C, Petersen MC, Hamner A. Swallowing function and medical diagnoses in infants suspected of Dysphagia. Pediatrics. 2001;108(6).
 53. Kakodkar K, Schroeder JW. Pediatric dysphagia. Pediatric Clinics of North America. 2013;60(4):969-77.
 54. Weir K, McMahon S, Barry L, Masters IB, Chang AB. Clinical signs and symptoms of oropharyngeal aspiration and dysphagia in children. The European respiratory journal : official journal of the European Society for Clinical Respiratory Physiology. 2009;33(3):604-11. Epub 2008/11/18.
 55. Richter GT. Management of oropharyngeal dysphagia in the neurologically intact and developmentally normal child. Current Opinion Otolaryngology Head Neck Surgery. 2010;18(6):554-63.
 56. Martin HB, Jones B. The videofluorographic swallowing study. Otorinolaryngology. 2009;59(1):19-29.
 57. RaberDurlacher JE, Brennan MT, Verdonck-de Leeuw IM, Cubson RJ, Eilers JG, Waltino T, et al. Swallowing dysfunction in cancer patients. Support Care Cancer. 2012;20:433-43.
 58. Cichero JAY, Hay GB, Murdoch BE, Halley PJ. Videofluoroscopic fluids versus mealtime fluids: differences in viscosity and density made clear. Journal of Medical Speech Language Pathology. 1997;5(3):203-15.
 59. DeMatteo C, Matovich D, Hjartarson A. Comparison of clinical and videofluoroscopic evaluation of children with feeding and swallowing difficulties. Developmental medicine and child neurology. 2005;47(3):149-57.

-
60. Ramsey DJC, Smithard DG, Kalra L. Early assessments of dysphagia and aspiration risk in acute stroke patients. *Stroke*. 2003;34(5):1252-7.
61. McCullough GH, Wertz RT, Rosenbek JC. Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. *Journal of communication disorders*. 2001;34(1-2):55-72. Epub 2001/04/27.
62. Cannon WB, Moser A. The movements of food in the esophagus. *American Journal of Physiology*. 1898;1:435-44.
63. Ramsey GH, Watson JS, Gramiak R, Weinberg SA. Cinefluorographic analysis of the mechanisms of swallowing. *Radiology*. 1955;64(4):498-518.
64. Palmer JB, Kuhlemeier KV, Tippet DC, Lynch C. A protocol for the videofluorographic swallowing study. *Dysphagia*. 1993;8(3):209-14.
65. Royal Rehabilitation Centre Sydney. Modified Barium Swallow Client Information Sheet. 2006 [cited 2013 cited 2013 17/9/2013]; Available from: Available from www.royalrehab.com.au/pdf/%20Documents/BariumSwallow.pdf.
66. The American College of Radiology. ACR practice guideline for the performance of the modified barium swallow. 2011 [cited 2013 cited 2013 17/9/2013]; Available from: <http://www.acr.org/~media/7d306289d61341dd9146466186a77dbe.pdf>.
67. Cichero J, Nicholson T, Dodrill P. Liquid barium is not representative of infant formula: characterisation of rheological and material properties. *Dysphagia*. 2011;26(3):264-71.
68. Mettler Jr FA, Huda W, Yoshizumi TT, Mahesh M. Effective doses in radiology and diagnostic nuclear medicine: A catalog. *Radiology*. 2008;248(1):254-63.
69. Weir KA, McMahon SM, Long G, Bunch JA, Pandeya N, Coakley KS, et al. Radiation doses to children during modified barium swallow studies. *Pediatric radiology*. 2007;37(3):283-90.
70. Bonilha HS, Humphries K, Blair J, Hill EG, McGrattan K, Carnes B, et al. Radiation exposure time during MBSS: Influence of swallowing impairment severity, medical diagnosis, clinician experience, and standardized protocol use. *Dysphagia*. 2013;28(1):77-85.
71. Mari F, Matei M, Ceravolo MG, Pisani A, Montesi A, Provinciali L. Predictive value of clinical indices in detecting aspiration in patients with neurological disorders. *Journal of neurology, neurosurgery, and psychiatry*. 1997;63(4):456-60. Epub 1997/10/29.
72. Warms T, Richards J. 'Wet voice' as a predictor of penetration and aspiration in oropharyngeal dysphagia. *Dysphagia*. 2000;15(2):84-8.
73. Arvedson J, Rogers B, Buck G, Smart P, Msall M. Silent aspiration prominent in children with dysphagia. *International Journal of Pediatric Otorhinolaryngology*. 1994;28(2-3):173-81.

-
74. Suiter DM, Leder SB. Clinical utility of the 3-ounce water swallow test. *Dysphagia*. 2008;23(3):321-37.
75. Romano M, Schultz T, Tai A. The Diagnostic Test Accuracy of Clinical Swallow Assessment for Oropharyngeal Aspiration: A Systematic Review 2012.
76. Bosma JF. Development of feeding. *Clinical Nutrition*. 1986;5(5):210-8.
77. American Speech Language Hearing Association. Roles of speech-language pathologists in swallowing and feeding disorders: technical report. USA: 2001.
78. Rutter MJ. Evaluation and management of upper airway disorders in children. *Seminars in Pediatric Surgery*. 2006;15(2):116-23.
79. Ruark JL, McCullough GH, Peters RL, Moore CA. Bolus consistency and swallowing in children and adults. *Dysphagia*. 2002;17(1):24-33.
80. Garon BR, Engle M, Ormiston C. Silent aspiration: results of 1,000 videofluoroscopic swallow evaluations. *Journal of Neurologic Rehabilitation*. 1996;10(2):121-6.
81. Gaziano JE. Evaluation and management of oropharyngeal dysphagia in head and neck cancer. *Cancer Control*. 2002;9(5):400-9.
82. The Joanna Briggs Institute. Quantitative Evidence. Joanna Briggs Institute reviewers' manual: 2011 edition. 2011 ed. Australia: The Joanna Briggs Institute; 2011. p. 45-76.
83. Whiting P, Rutjes AWS, Reitsma JB, Bossuyt PMM, Kleijrien J. The development of QUADAS: a tool for the quality assessment of studies of diagnostic test accuracy included in systematic reviews. *BMC Medical Research Methodology*. 2003;3(25):1-13.
84. Meyer GJ. Guidelines for reporting information in studies of diagnostic test accuracy: the STARD initiative. *Journal of Personality Assessment*. 2003;81(3):191-3.
85. Field MJ, Behrman RE. Ethical Conduct of Clinical Research Involving Children. Washington (DC): National Academies Press (US); 2004; Available from: <http://www.ncbi.nlm.nih.gov/books/NBK25557/>.
86. Dinnes J, Deeks J, Kirby J, Roderick P. A methodological review of how heterogeneity has been examined in systematic reviews of diagnostic test accuracy. *Health Technology Assessment*. 2005;9(12):1 - 113.
87. Bours GJJ, Speyer R, Lemmens J, Limburg M, de Wit R. Bedside screening tests vs. videofluoroscopy or fiberoptic endoscopic evaluation of swallowing to detect dysphagia in patients with neurological disorders: systematic review. *Journal of Advanced Nursing*. 2009;65(3):477-93.
88. Barbiera F, Condello S, De Palo A, Todaro D, Mandracchia C, De Cicco D. Role of videofluorography swallow study in management of dysphagia in neurologically compromised patients. *La Radiologia medica*. 2006;111(6):818-27. Epub 2006/08/10.

-
89. Linden P, Kuhlemeier KV, Patterson C. The probability of correctly predicting subglottic penetration from clinical observations. *Dysphagia*. 1993;8(3):170-9.
90. Baylow HE, Goldfarb R, Taveira CH, Steinberg RS. Accuracy of clinical judgment of the chin-down posture for dysphagia during the clinical/bedside assessment as corroborated by videofluoroscopy in adults with acute stroke. *Dysphagia*. 2009;24(4):423-33.
91. Splaingard ML, Hutchins B, Sulton LD, Chaudhuri G. Aspiration in rehabilitation patients: videofluoroscopy vs bedside clinical assessment. *Archives of physical medicine and rehabilitation*. 1988;69(8):637-40. Epub 1988/08/01.
92. Zhou Z, Salle JY, Daviet JC, Stuit A, Nguyen CL. Combined approach in bedside assessment of aspiration risk post stroke: PASS. *European Journal of Physical and Rehabilitation Medicine*. 2011;47(3):441-6.
93. McCullough GH, Rosenbek JC, Wertz RT, McCoy S, Mann G, McCullough K. Utility of clinical swallowing examination measures for detecting aspiration post-stroke. *Journal of Speech, Language, and Hearing Research*. 2005;48(6):1280-93.
94. Smithard DG, O'Neill PA, Park C, England R, Renwick DS, Wyatt R, et al. Can bedside assessment reliably exclude aspiration following acute stroke? *Age and Ageing*. 1998;27(2):99-106.
95. Mann G, Hankey GJ, Cameron D. Swallowing disorders following acute stroke: Prevalence and diagnostic accuracy. *Cerebrovascular Diseases*. 2000;10(5):380-6.
96. Hammond CAS, Goldstein LB, Horner RD, Ying J, Gray L, Gonzalez-Rothi L, et al. Predicting aspiration in patients with ischemic stroke. *CHEST*. 2009;135(3):769-77.
97. DePippo KL, Holas MA, Reding MJ. Validation of the 3-oz water swallow test for aspiration following stroke. *Archives of Neurology*. 1992;49(12):1259-61.
98. Tohara H, Saitoh E, Mays KA, Kuhlemeier K, Palmer JB. Three tests for predicting aspiration without videofluorography. *Dysphagia*. 2003;18(2):126-34.
99. Barbiera F, Fiorentino E, D'Agostino T, Acquaro P, Ferraro L, Parisi M, et al. Digital cineradiographic swallow study: our experience. *La Radiologia medica*. 2002;104(3):125-33. Epub 2002/12/10.
100. Linden P, Siebens AA. Dysphagia: Predicting laryngeal penetration. *Archives of Physical Medicine and Rehabilitation*. 1983;64(6):281-4.
101. Schwartz A. Diagnostic Test Calculator. [cited 2013 4/8]; Available from: araw.mede.uic.edu/cgi-bin/testcalc.pl.
102. Macaskill P, Gatsonis C, Deeks J, Harbord R. Chapter 10: Analysing and Presenting Results. *The Cochrane Collaboration 2010* [cited 2013 10/08]; Available from: <http://srdta.cochrane.org/>.

-
103. Reitsma JB, Glas AS, Rutjes AWS, Scholten RJPM, Bossuyt PM, Zwinderman AH. Bivariate analysis of sensitivity and specificity produces informative summary measures in diagnostic reviews. *Journal of Clinical Epidemiology*. 2005;58(10):982-90.
104. Terre R, Mearin F. Oropharyngeal dysphagia after the acute phase of stroke: predictors of aspiration. *Neurogastroenterology and Motility: the official journal of the European Gastrointestinal Motility Society*. 2006;18(3):200-5.
105. Garon BR, Sierzant T, Ormiston C. Silent aspiration: results of 2,000 video fluoroscopic evaluations. *Journal of Neuroscience Nursing*. 2009;41(4):178-85.
106. Weir KA, McMahon S, Taylor S, Chang AB. Oropharyngeal aspiration and silent aspiration in children. *CHEST*. 2011;140(3):589-97.
107. Cass H, Wallis C, Ryan M, Reilly S, McHugh K. Assessing pulmonary consequences of dysphagia in children with neurological disabilities: when to intervene? *Developmental Medicine and Child Neurology*. 2005;47(5):347-52.
108. Park YH, Han HR, Oh BM, Lee J, Park JA, Yu SJ, et al. Prevalence and associated factors of dysphagia in nursing home residents. *Geriatric Nursing*. 2013;34(3):212-7.
109. Chen MY, Peele VN, Donati D, Ott DJ, Donofrio PD, Gelfand DW. Clinical and videofluoroscopic evaluation of swallowing in 41 patients with neurologic disease. *Gastrointest Radiol*. 1992;17(2):95-8. Epub 1992/01/01.
110. Clavé P, Arreola V, Romea M, Medina L, Palomera E, Serra-Prat M. Accuracy of the volume-viscosity swallow test for clinical screening of oropharyngeal dysphagia and aspiration. *Clinical Nutrition*. 2008;27(6):806-15.
111. Daniels SK, Brailey K, Priestly DH, Herrington LR, Weisberg LA, Foundas AL. Aspiration in patients with acute stroke. *Archives of physical medicine and rehabilitation*. 1998;79(1):14-9. Epub 1998/01/24.
112. Daniels SK, McAdam CP, Brailey K, Foundas AL. Clinical assessment of swallowing and prediction of dysphagia severity. *American Journal of Speech-Language Pathology*. 1997;6(4):17-24.
113. Groher ME, Crary MA, Carnaby G, Vickers Z, Aguilar C. The impact of rheologically controlled materials on the identification of airway compromise on the clinical and videofluoroscopic swallowing examinations. *Dysphagia (0179051X)*. 2006;21(4):218-25.
114. Horiguchi S, Suzuki Y. Screening tests in evaluating swallowing function. *Japan Medical Association Journal*. 2011;54(1):31-4.
115. Kidd D, Lawson J, Nesbitt R, MacMahon J. Aspiration in acute stroke: a clinical study with videofluoroscopy. *The Quarterly journal of medicine*. 1993;86(12):825-9. Epub 1993/12/01.

-
116. Kopey SA, Chae J, Vargo MM. Does a 3-sip test detect dysphagia in acute stroke rehabilitation patients? *Dysphagia*. 2011;26(2):202-3.
117. Leslie P, Drinnan MJ, Finn P, Ford GA, Wilson JA. Reliability and validity of cervical auscultation: A controlled comparison using videofluoroscopy. *Dysphagia*. 2004;19(4):231-40.
118. Liesching J. Is clinical evaluation enough? A critique of the clinical bedside examination in dysphagia. *Asia Pacific Journal of Speech, Language & Hearing*. 2003;8(2):81-5.
119. Marrara JL, Duca AP, Dantas RO, Trawitzki LV, Lima RA, Pereira JC. Swallowing in children with neurologic disorders: clinical and videofluoroscopic evaluations. *Pró-fono : revista de atualização científica*. 2008;20(4):231-6.
120. Martino R, Pron G, Diamant N. Screening for oropharyngeal dysphagia in stroke: Insufficient evidence for guidelines. *Dysphagia*. 2000;15(1):19-30.
121. McCann LJ, Garay SM, Ryan MM, Harris R, Riley P, Pilkington CA. Oropharyngeal dysphagia in juvenile dermatomyositis (JDM): An evaluation of videofluoroscopy swallow study (VFSS) changes in relation to clinical symptoms and objective muscle scores. *Rheumatology*. 2007;46(8):1363-6.
122. Nishiwaki K, Tsuji T, Liu M, Hase K, Tanaka N, Fujiwara T. Identification of a simple screening tool for dysphagia in patients with stroke using factor analysis of multiple dysphagia variables. *Journal of Rehabilitation Medicine (Taylor & Francis Ltd)*. 2005;37(4):247-51.
123. Perry L, Love CP. Screening for dysphagia and aspiration in acute stroke: A systematic review. *Dysphagia*. 2001;16(1):7-18.
124. Rosenbek JC, McCullough GH, Wertz RT. Is the information about a test important? Applying the methods of evidence-based medicine to the clinical examination of swallowing. *Journal of Communication Disorders*. 2004;37(5):437-50.
125. Shem KL, Castillo K, Wong SL, Chang J, Kao MC, Kolakowsky-Hayner SA. Diagnostic Accuracy of Bedside Swallow Evaluation Versus Videofluoroscopy to Assess Dysphagia in Individuals With Tetraplegia. *PM and R*. 2012;4(4):283-9.
126. Silva AC, Dantas RO, Fabio SR. Clinical and scintigraphic swallowing evaluation of post-stroke patients. *Pro-fono : revista de atualizacao científica*. 2010;22(3):317-24. Epub 2010/11/26.
127. Waito A, Bailey GL, Molfenter SM, Zoratto DC, Steele CM. Voice-quality abnormalities as a sign of dysphagia: validation against acoustic and videofluoroscopic data. *Dysphagia*. 2011;26(2):125-34. Epub 2010/05/11.
128. Wu MC, Chang YC, Wang TG, Lin LC. Evaluating Swallowing Dysfunction Using a 100-ml Water Swallowing Test. *Dysphagia*. 2004;19(1):43-7.

129. Zenner PM, Losinski DS, Mills RH. Using cervical auscultation in the clinical dysphagia examination in long-term care. *Dysphagia*. 1995;10(1):27-31.

8 Appendices

8.1 Appendix I: Initial Search Terms

deglutition

deglutition disorders

oesophageal motility disorders

swallowing

speech therapy

physical examination

neurologic examination

fluoroscopy

videofluoroscopy

videofluorography

photofluorography

aspiration

pneumonia, aspiration

respiratory aspiration

sensitivity

specificity

diagnostic test

accuracy

predictive value

8.2 Appendix II: Electronic databases

PubMed

EMBASE

CINAHL

ERIC

Scopus

Cochrane Library

Web of Science

Web of Knowledge

Mednar

EthOS

ProQuest

Networked Digital Library of Theses and Dissertations

DART-Europe E-theses portal

8.3 Appendix III: PubMed Search Terms

deglutition disorders[mh:noexp] OR esophageal motility disorders[mh:noexp] OR esophageal motility[tw] OR oesophageal motility[tw] OR deglutition[mh] OR deglutition[tw] OR swallowing[tw] OR dysphagia[tw] OR deglutition disorder[tw] OR oropharyngeal dysphagia[tw] OR pharynx[mh]	physical examination[mh:noexp] OR neurologic examination[mh:noexp] OR neurologic exam*[tw] OR neurological exam*[tw] OR assess*[tw] OR swallow[tw] OR clinical evaluation[tw] OR bedside evaluation[tw] OR cervical auscultation[tw] OR speech pathol*[tw] OR speech language therapist[tw] OR speech-language therapist[tw] OR speech therap*[tw] OR occupational therap*[tw] OR speech therapy[mh] OR occupational therapy[mh] OR penetration aspiration scale[tw]	fluoroscopy[mh] OR fluoroscop*[tw] OR fluroscop*[tw] OR photofluorog*[tw] OR photoflurog*[tw] OR videofluorog*[tw] OR recording[mh] OR recording*[tw] OR videorecording [tw] OR videoendoscop*[tw] OR videofluoro*[tw] OR cineradiography[mh:noexp] OR cineradiog*[tw] OR cinefluorog*[tw] OR cineflurog*[tw] OR scintigraph*[tw]	OR respiratory aspiration[mh] OR aspiration[tw] OR pneumonia, aspiration[mh] OR mendelson syndrome[tw] OR mendelson's syndrome[tw] OR inhalation[mh] OR inhal*[tw] OR cough[mh] OR cough[tw] OR respirat*[tw]	diagnosis[mh:noexp] OR sensitivity and specificity[mh] OR sensitiv*[tw] OR diagnos*[tw] OR OR diagnostic[mh:noexp] OR diagnosis, differential[mh:noexp] OR diagnosis[sh:noexp]
--	--	---	---	--

8.4 Appendix IV: Critical appraisal tool – the QUADAS check list ⁸³

1. Was the spectrum of patients representative of the patients who will receive the test in practice?
2. Were selection criteria clearly described?
3. Is the reference standard likely to correctly classify the target condition?
4. Is the time period between reference standard and index test short enough to be reasonably sure that the target condition did not change between the two tests?
5. Did the whole sample or a random selection of the sample, receive verification using a reference standard of diagnosis?
6. Did patients receive the same reference standard regardless of the index test result?
7. Was the reference standard independent of the index test (i.e. the index test did not form part of the reference standard)?
8. Was the execution of the index test described in sufficient detail to permit replication of the test?
9. Was the execution of the reference standard described in sufficient detail to permit its replication?
10. Were the index test results interpreted without knowledge of the results of the reference standard?
11. Were the reference standard results interpreted without knowledge of the results of the index test?
12. Were the same clinical data available when test results were interpreted as would be available when the test is used in practice?
13. Were uninterpretable/ intermediate test results reported?
14. Were withdrawals from the study explained?

8.5 Appendix V: Data extraction tool – the STARD check list ⁸⁴

- 1 Was the study identified as being a diagnostic accuracy study?
- 2 Were research questions or study aims, such as estimating diagnostic accuracy or comparing accuracy between tests or across participant groups, detailed?
- 3 Does the study describe the study population, inclusion and exclusion criteria, setting and locations where the data were collected?
- 4 Does the study describe participant recruitment? Was recruitment based on presenting symptoms, results from previous tests, or the fact that the participants had received the index tests or the reference standard?
- 5 Describe participant sampling: Was the study population a consecutive series of participants defined by the selection criteria in items 3 and 4? If not, specify how participants were further selected
- 6 Describe data collection: Was data collection planned before the index test and reference standard were performed (prospective study) or after (retrospective study)?
- 7 Did the study describe the reference standard and its rationale?
- 8 Describe technical specifications of material and methods involved including how and when measurements were taken, and/or cite references for index tests and reference standard?
- 9 Describe definition of and rationale for the units, cut-offs and/or categories of the results of the index tests and the reference standard.
- 10 Describe the number, training and expertise of the persons executing and reading the index tests and the reference standard
- 11 Describe whether or not the readers of the index tests and reference standard were blind (masked) to the results of the other test and describe any other clinical information available to the readers.

12 Describe methods for calculating or comparing measures of diagnostic accuracy, and the statistical methods used to quantify uncertainty (e.g. 95% confidence intervals)

13 Describe methods for calculating test reproducibility, if done

14 Report when study was done, including beginning and ending dates of recruitment

15 Does the study report clinical and demographic characteristics of the study population (e.g. age, sex, spectrum of presenting symptoms, co morbidity, current treatments, recruitment centers)?

16 Does the study report the number of participants satisfying the criteria for inclusion that did or did not undergo the index tests and/or the reference standard; describe why participants failed to receive either test (a flow diagram is strongly recommended) ?

17 Does the study report time interval from the index tests to the reference standard, and any treatment administered between?

18 Does the study report distribution of severity of disease (define criteria) in those with the target condition; other diagnoses in participants without the target condition?

19 Does the study report a cross tabulation of the results of the index tests (including indeterminate and missing results) by the results of the reference standard; for continuous results, the distribution of the test results by the results of the reference standard?

20 Does the study report any adverse events from performing the index tests or the reference standard?

21 Does the study report estimates of diagnostic accuracy and measures of statistical uncertainty (e.g. 95% confidence intervals)?

22 Does the study report how indeterminate results, missing responses and outliers of the index tests were handled?

23 Does the study report estimates of variability of diagnostic accuracy between subgroups of participants, readers or centers?

24 Does the study report estimates of test reproducibility?

25 Does the study discuss the clinical applicability of the study findings?

8.6 Appendix VI: Data extracted from individual studies and combined in meta-analysis

Study, year.	TP	FP	FN	TN
Barbiera et al. 2006	17	12	4	14
Baylow et al. 2009	2	5	3	20
DeMatteo et al. 2005	22	19	2	16
DePippo et al. 1992	16	11	4	13
Hammond et al. 2009	19	11	14	52
Linden et al. 1993	64	46	36	92
Mann et al. 2000	26	37	2	63
McCullough et al. 2001	17	14	5	24
McCullough et al. 2005	23	13	20	109
Smithard et al. 1998	9	9	10	55
Splaingard et al. 1988	6	4	23	52
Tohara et al. 2003	20	4	9	30
Zhou et al. 2011	48	10	6	43

TP – true positive; FP- false positive; FN – false negative; TN – true negative

8.7 Appendix VII: Table of included studies

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
Barbiera et al. 2006	47 participants. Neurological deficits including cerebrovascular accident (CVA), head-brain trauma. Age 16 – 80 years. Examined within 6 – 12 months onset of dysphagia (not onset disease diagnosis).	Video Fluoroscopic Swallow Study, performed by Radiologist. Radiologist unaware of results of index test.	Complete swallowing assessment, does not state which profession performed the assessment. Scored as – dysphagic with signs aspiration, dysphagic no signs aspiration, non dysphagic.	Unclear	Not stated.	81%	54%	QUADAS score 8: medium risk

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
Baylow et al. 2009	15 adult acute stroke patients – haemorrhagic or ischaemic. Mean age 64;6 (48-80yrs); mean days post onset cerebrovascular accident (CVA) 9.2 (1-35)	Video Fluoroscopic Swallow Study (VFSS) – rated by speech pathologist and radiologist as aspiration present or absent.	Performed by speech pathologist. Pre specified protocol used. 2 x swallow trials of each bolus per patient. Pre specified bolus size. Same bolus size used for VFSS.	Solids and fluids provided. This systematic review provides results for thin fluids only from this study.	VFSS within 96 hrs of clinical swallow assessment (CSA).	40%	80%	QUADAS 10 - medium
DeMatteo et al. 2005	59 participants. Infants and children; 0-14 years; 62 % younger than 12 months. Diagnoses include: cerebral palsy, prematurity, Pierre Robin, hypoxic-ischaemic encephalopathy,	Video fluoroscopic swallow study performed by a different therapist from the clinical swallow assessment; films confirmed with Radiologist.	Complete clinical swallow evaluation with a checklist provided and reference for use of this assessment. Scoring: Therapists' level of confidence about suspicion of aspiration: 100% = certainty aspiration present. 0% = certainty	Fluids	Same day or within 48 hours.	92%	46%	QUADAS score 12 – low risk

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
	vacterl syndrome, angelman syndrome, infantile spasms, cardiac, downs', developmental delay, seizures, failure to thrive, acquired brain injury, brain tumour		aspiration not present. 50% = uncertainty					
De Pippo et al. 1992	44 participants. Adults (71 +/- 10 years) Diagnosis stroke (5 +/- 3 weeks post onset)	Modified Barium Swallow (equivalent to a Video Fluoroscopic Swallow Study) Reviewed by two Speech Pathologists	Screening tool 3-oz water, drink consecutively from cup. Scoring: Swallow recorded as 'abnormal' if coughing during or one minute after, or presence of post swallow wet-hoarse voice quality	Fluids	Information not provided	80%	54%	QUADAS score 10 – medium risk.

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
Hammond et al. 2009	96 adults (mean age 67, standard deviation 1.19); recent ischaemic CVA (time post onset of symptoms not stated). Exclusion criteria – head and neck cancer; brain tumour; neurosurgery.	Video Fluoroscopic Swallow Study (n=91); FEES (fiberoptic endoscopic evaluation of swallow) (n=5). Not stated the profession who performed the reference standard test, however films were recorded and evaluated by speech pathologist blind to results of index test.	Performed by speech pathologists, unaware of results of reference standard. Included assessment of swallow integrity and presence of cough after fluid trial.	Fluids	Index test performed immediately before or after reference standard.	58%	83%	QUADAS score 11 – low risk

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
Linden et al. 1993	249 patients, results for 238 patients provided, unclear remaining missing data. 95% of patients were adults with neurological aetiology as cause for dysphagia, most common – stroke. Remainder not stated.	Video Fluoroscopic Swallow Study (VFSS) after the index test. Protocol used for VFSS, no other details provided.	Administered by a Speech Pathologist 'knowledgeable about dysphagia and its treatment' – protocol for the assessment provided as appendix.	Not specified.	Not specified.	64%	67%	QUADAS – 5, high risk.
Mann et al. 2000	128 adult patients, mean age 71 years (SD 12.1 years). Acute stroke (< 7 days since symptom onset).	Modified barium swallow performed by radiologist. Films reviewed by radiologist and a speech pathologist. 5 point scale ranging from 'no aspiration' to 'frank	Performed by two speech pathologists blind to video fluoroscopy findings using a standard protocol (provided as an appendices).	Fluids.	Unclear. Authors state the index test was performed within 3 days of symptom onset and video	93%	63%	QUADAS score 12 – low risk

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
		aspiration'.			fluoroscopy within 10 days.			
McCullough et al. 2001	60 participants. Adults, mean age 67.8 (range 40-96). Diagnosed stroke within 6 weeks. Mean days since diagnosis of stroke 5.98 (range 1-42).	Video Fluoroscopic Swallow Study (VFSS). VFSS films reviewed and rated for aspiration (present or absent) at least 1 week after completion of the VFSS blinded to results of clinical swallow assessment.	Complete clinical swallow assessment using four sections – history, oral motor, voice and swallow trials. Scored with a binary system – overall estimate of presence or absence of aspiration.	Fluids and solids combined.	Within 24 hours	77%	63%	QUADAS score 12 – low risk.
McCullough et al. 2005	165 participants. Adults mean age 65 years (39-101 years). Acute ischaemic stroke, occurrence	Video Fluoroscopic Swallow Study (VFSS). VFSS films reviewed and rated for aspiration	Complete clinical swallow assessment using four sections – history, oral motor, voice and swallow trials.	Fluids and solids reported separately.	Within 24 hours	53%	89%	QUADAS score 12 low risk

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
	within 6 weeks of examination.	(present or absent) at least 1 week after completion of the VFSS blinded to results of clinical swallow assessment.	Scored with a binary system – overall estimate of presence or absence of aspiration.					
Smithard et al. 1998	83 participants (assessed by Speech-Language Therapist 1). Adults presenting within 24 hours of the onset of acute stroke, over a 12 month period. Age 40-93,	Video Fluoroscopic Swallow Study - performed within 3 days of the stroke.	Results are presented in this review for Speech-Language Pathologist (SLT) 1 only. Clinical swallow assessment within 24 hours of reference test. Standardised bedside swallow assessment used.	Fluids only	24 hours	47%	86%	QUADAS 11 – low risk

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
	median age 79.							
Splaingard et al. 1988	85 participants. Adults and children (less than 15 years); range of diagnoses including stroke and brain injury. 23 patients had a tracheostomy, naso gastric tube and/or PEG. Results presented separately for patients with and without a tracheostomy, NG and/or PEG. Does not separate results of adults from children (10 children); did not state mean time	Video Fluoroscopic Swallow Study – standard protocol. Performed by speech pathologist and physician.	Complete clinical swallow assessment performed by 'certified' speech pathologist including case history, oral examination and swallowing trials. Score out of 5 , Score of 4 or 5 = aspiration.	Does not separate results for fluids from solids.	Reference standard within 72 hours after index test.	21%	93%	QUADAS score 11 – low risk.

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
	post onset of cause for dysphagia. Referred by physician for possible swallowing dysfunction.							
Tohara et al. 2003	63 participants. Adults with symptoms of dysphagia. Age range 18-83 yrs. Mean age 63 years, standard deviation +/- 17 years. Diagnoses: 36 stroke, 7 neuropathy or myopathy, 6 traumatic encephalopathy, 5 respiratory disease, 4 brain tumour, 2 oral or	Performed by rehabilitation physician and/or dentist trained in dysphagia evaluation. Did not use motion x-ray – still x-ray taken prior to swallowing thin barium solution and after the swallow. Films reviewed for presence of aspiration.	Performed by rehabilitation physician and/or dentist trained in dysphagia evaluation. Water test only not complete swallow assessment.	Fluids only.	Reference test performed one week after index test.	69%	88%	QUADAS score 9 – medium risk.

Study Citation	Population	Reference Test Details	Index Test Details	Results for fluids and / or solids	Time between Index Test and Reference Test	Sensitivity	Specificity	Risk of Bias
	pharyngeal tumour, 1 spinal cord injury, 1 'inamination', 1 unknown aetiology.							
Zhou et al. 2011	107 consecutive patients with first ever stroke (average age 67 +/- 13 years), mean time after stroke onset 7 +/- 5 months. Exclusion criteria – history of otolaryngological cancer, below 18 years of age.	Video Fluoroscopic Swallow Study performed by two rehabilitation physicians specialised in this examination, interpreted blindly to results of index test.	Performed by two physicians (1 neurologist, 1 rehabilitation physician). Practical Aspiration Screening Schema (PASS): use of 90 mls water swallow test - drinking 90mls water without interruption. Scored yes for aspiration present if: task not completed, coughing, choking, wet-hoarse voice during or within one minute after end of test.	Fluids and solids.	Not specified.	89%	81%	QUADAS score 11 – low risk.

8.8 Appendix VIII: Table of excluded studies

Reference	Reason for exclusion
Arvedson et al. 1994 ⁷³	Review article, not a study of diagnostic test accuracy.
Bours et al. 2009 ⁸⁷	Systematic review, not a single study of diagnostic test accuracy.
Chen et al. 1992 ¹⁰⁹	Not a study of diagnostic test accuracy.
Clave et al. 2008 ¹¹⁰	Almost one third of participants have diagnosis of head and neck cancer (exclusion criteria for this systematic review); use of pulse oximeter during clinical swallow assessment which does not satisfy definition of clinical swallow assessment for this systematic review as it is not a standard item used for clinical swallow assessment in clinical practice.
Daniels et al. 1998 ¹¹¹	Provides sensitivity and specificity for various features of the clinical swallow assessment in predicting aspiration. A binary scoring system is not used for the clinical swallow assessment, the clinician performing the clinical swallow assessment does not state whether or not their assessment findings demonstrate presence or absence of aspiration.
Daniels et al. 1997 ¹¹²	Diagnosis is for dysphagia severity not oropharyngeal aspiration.
Groher et al. 2006 ¹¹³	Provides sensitivity and specificity for cough and/or wet voice in predicting aspiration. A binary scoring system is not used for the clinical swallow assessment. The clinician performing the clinical swallow assessment does not state whether or not their assessment findings

	demonstrate presence or absence of aspiration.
Horiguchi et al. 2011 ¹¹⁴	Not a study of diagnostic test accuracy.
Kidd et al. 1993 ¹¹⁵	Paper lists each feature of clinical swallow assessment and compares number of participants with aspiration present or absent on video fluoroscopy.
Kopey et al. 2011 ¹¹⁶	Diagnosis is not for oropharyngeal aspiration, it is for 'clinically relevant dysphagia'. The definition does not match the definition of oropharyngeal aspiration described in this systematic review paper.
Leslie et al. 2004 ¹¹⁷	The features of this clinical swallow assessment do not meet the specified criteria for clinical swallow assessment in this systematic review. The clinical swallow assessment only uses cervical auscultation and involves recording the sound clips.
Liesching et al. 2003 ¹¹⁸	Review article, not a study of diagnostic test accuracy.
Mari et al. 1997 ⁷¹	Provides sensitivity and specificity for a specific set of features of the clinical swallow assessment. A binary scoring system is not used for the clinical swallow assessment. The clinician performing the clinical swallow assessment does not state whether or not their assessment findings demonstrate presence or absence of aspiration.
Marrara et al. 2008 ¹¹⁹	Not a study of diagnostic test accuracy. Presents the correlation between findings of the clinical swallow assessment and video fluoroscopic swallow study.
Martino et	Review article presenting findings of multiple studies, not a single

al. 2000 ¹²⁰	study of diagnostic test accuracy.
McCann et al. 2007 ¹²¹	Not a study of diagnostic test accuracy. Presents the correlation between findings of the clinical swallow assessment and video fluoroscopic swallow study.
Nishiwaki et al. 2005 ¹²²	Provides sensitivity and specificity for various features of the clinical swallow assessment in predicting aspiration. A binary scoring system is not used for the clinical swallow assessment, the clinician performing the clinical swallow assessment does not state whether or not their assessment findings demonstrate presence or absence of aspiration.
Perry et al. 2001 ¹²³	Systematic review, not a single study of diagnostic test accuracy.
Rosenbek et al. 2004 ¹²⁴	Provides sensitivity and specificity for various features of the clinical swallow assessment in predicting aspiration. A binary scoring system is not used for the clinical swallow assessment, the clinician performing the clinical swallow assessment does not state whether or not their assessment findings demonstrate presence or absence of aspiration.
Shem et al. 2012 ¹²⁵	Diagnosis is not for oropharyngeal aspiration, it is for dysphagia.
Silva et al. 2009 ¹²⁶	Descriptive study, not a study of diagnostic test accuracy.
Waito et al. 2011 ¹²⁷	Does not compare clinical swallow assessment with video fluoroscopic swallow study, compares acoustic measures of voice quality.
Wu et al. 2004 ¹²⁸	Compares episodes of choking during water swallow test with findings of aspiration on the video fluoroscopic swallow study. A binary scoring system is not used for the clinical swallow assessment. The clinician performing the clinical swallow assessment does not state whether or not their assessment findings demonstrate presence or absence of

aspiration.

Zenner et al. 1995 ¹²⁹ Unable to obtain data from results, presented in narrative form only.
The author was not able to be contacted via email.

