

Inter- and intra-rater reliability of cervical auscultation to detect aspiration in patients with dysphagia

AE Stroud Department of Medicine (Neurology), University of Wales College of Medicine, **BW Lawrie** Department of Radiology, University Hospital of Wales and **CM Wiles** Department of Medicine (Neurology), University of Wales College of Medicine, Cardiff, Wales, UK

Received 22nd January 2001; returned for revisions 1st April 2001; revised manuscript accepted 25th May 2001.

Objective: To measure the inter- and intra-rater reliability of cervical auscultation used alone to detect aspiration in dysphagic patients.

Setting: A university teaching hospital.

Design: Comparison of the detection of aspiration in 16 recorded swallow sounds by five speech and language therapists on two occasions. Swallow sounds were recorded simultaneously with videofluoroscopy.

Subjects: Sixteen patients referred for assessment of dysphagia with videofluoroscopy.

Results: The kappa statistic for multiple raters showed fair agreement between raters ($\kappa = 0.28$). There was high agreement when aspiration occurred but in non-aspirating swallows there was significant over-detection of aspiration ($p < 0.001$ McNemar's test). The intra-rater reliability within different individuals was widely variable ($\kappa = 0.55$ (range 0.31–0.85)).

Conclusions: Presented with the swallowing sounds in isolation speech and language therapists cannot reliably classify swallows into those with accompanying aspiration and those without. There appears to be a problem of over-detection of aspiration. Even in this small study, however, some individual therapists achieve such high reliability ($\kappa = 0.85$) that they must be using successful internal criteria to interpret the swallow sounds correctly and further qualitative research may identify these.

Introduction

Swallowing difficulties arise from numerous aetiologies.¹ Forty-five per cent of stroke patients suffer from dysphagia,² which increases morbidity

and mortality and is a risk factor for aspiration pneumonia, upper airway obstruction, malnutrition and dehydration.³ If detected, complications can be successfully avoided.

Reliability of clinical bedside measures for detection of aspiration has been found to be low with questionable utility.⁴ Videofluoroscopy is the current 'gold standard'⁵ with high reliability for detection of aspiration,⁶ but is expensive, inappropriate for aspiration screening and moni-

Address for correspondence: Alison Stroud, Speech and Language Therapist, Prince Charles Hospital, Merthyr Tydfil, Mid Glamorgan, CF47 9DT, UK. e-mail: Alison.Stroud@nglam-tr.wales.nhs.uk

toring, involves radiation and is physically demanding for some patients.

Cervical auscultation is a method of listening at the lateral aspect of the thyroid cartilage to the sounds of swallowing and respiration in order to detect aspiration; almost all normal people breathe out after a swallow.⁷ It has potential as a cheap, portable tool. Although previous studies suggest it has predictive validity with a reported sensitivity between 90% and 73% and specificity around 80%,^{8,9} test reliability is unknown. This study examines the reliability of cervical auscultation used in isolation from other clinical cues.

Swallowing sounds were recorded simultaneously with videofluoroscopy examination and played without the picture, blind, to experienced speech and language therapists. The videofluoroscopy recording was used to classify whether an aspiration event had occurred in any particular swallow. Thus this study measures reliability by comparing inter- and intra-rater detection of aspiration. It comments on concurrent criterion validity by comparing the videofluoroscopy classification of the swallow to the cervical auscultation classification.

Methods

Recruitment

Patients were consecutively recruited from those referred to the videofluoroscopy clinic as part of their normal clinical management for dysphagia, at a large teaching hospital. All patients received an information letter with their clinic appointment. On the day, verbal information was given, with an opportunity to ask questions and withdraw without prejudice to treatment, following which written consent was obtained. The general practitioner was informed. Ethical approval was obtained from Bro Taf Local Research and Ethics Committee.

Procedure

Simultaneous recordings of the swallow sounds and the videofluoroscopy picture were made utilizing a unique radiotranslucent stethoscope head containing an electret microphone onto a Panasonic S-VHS video recorder AG-7330. The

stethoscope head was placed on the patient's neck on the lateral aspect of the thyroid cartilage by the researcher and held in place by a 3M Double Stick Disc 2181. Optimum recording levels were set on hi-fi during a dry swallow. The standard hospital videofluoroscopy protocol was followed and bolus type and size were recorded. The decision on categorization of the swallow was made through discussion and agreement between the consultant radiologist and the clinical speech and language therapist, which has substantial reliability for detection of aspiration using videofluoroscopy ($\kappa = 0.80$).⁶ Aspiration was determined using standard criteria.¹⁰ To allow efficient playback for the reliability studies the recordings were edited on a Media 100 XS nonlinear edit suite.

Five speech and language therapists experienced in cervical auscultation were recruited from a hospital setting in Bro Taf Health Authority. Experience was defined as

- having received at least 5 hours training on cervical auscultation and
- having used auscultation to assess patients for at least one session per week as part of their current job.

Each therapist was asked to score aspiration categorically. The swallow sound clips were played in random order (by picking a number from a box) to each therapist, separated from each other, blind to the video picture, results and clinical data. The sounds were played through Sennheiser HD 40 headphones on a Panasonic S-VHS player AG 7130. This was repeated in re-randomized order, two weeks later giving no further information.

Statistical analysis

The kappa approach recommended by Fleiss¹¹ for agreement over chance between multiple raters was used. Kappa scores for intra-rater reliability in individuals were calculated using SPSS for Windows. The benchmarks provided by Landis and Koch¹² to rate kappa values on a scale of 'poor' to 'almost perfect', although familiar and popularly used, can be oversimplistic if regarded as universally applicable. Therefore results were also interpreted in context. Potential sources of

error were identified. Sensitivity, specificity, positive and negative predictive values were calculated from the first set of data from therapists, so the calculations were based on 80 recordings which came from five therapists each rating 16 swallows once. Actual yes/no response for aspiration detection was used, regardless of correctness according to videofluoroscopy.

Results

The 16 patients were four men and 12 women, mean age 51.8 (range 29.5–65 years). Eleven were being investigated for dysphagia of unknown origin, one had **pseudobulbar palsy** and four had had surgery for **head and neck cancer**. Average bolus size was 5 ml (2.5–10 ml). Three swallows resulted in aspiration and 13 in non-aspiration, which mimics the prevalence in other dysphagic populations.^{13,14}

Agreement

In considering the degree to which raters agreed from one occasion to the next and between each other it is useful to also know whether their interpretation was correct or incorrect. In Table 1 therefore the category of each swallow evaluation by cervical auscultation (aspiration or no aspiration) is shown for each rater on two occasions in a way which allows interpretation to be seen as correct (1) or incorrect (0).

The level of agreement over chance between the multiple raters was 0.28 (kappa), regardless of correctness according to videofluoroscopy. Overall this may be described as ‘fair’ agreement using Landis and Koch’s benchmarks. However, disagreements occurred much more in the context of non-aspirating swallows, rather than aspirating swallows. When aspiration occurred agreement was ‘almost perfect’, and accurate as judged by the gold standard videofluoroscopy: **thus therapists appeared to be able to detect true aspiration correctly and therefore agreed with each other. Reliability was reduced when aspiration did not occur; there was a significant occurrence of false positive results (McNemar’s test $p < 0.001$). Sensitivity was 86%, specificity 56%, positive predictive value 31% and negative predictive value 94%. There was a wide range**

of intra-rater reliability with kappa values ranging from 0.31 to 0.85, mean 0.55.

Previous studies have suggested cervical auscultation may detect aspiration by picking up the sound of air mixing with water in the airway after a swallow.¹⁵ Because of the tendency to report aspiration using auscultation when none was actually present the non-aspirating swallows were therefore examined in retrospect. In a patient with **pseudobulbar palsy** raters were all in **agreement but were consistently incorrect** (swallow 16). An unusual voicing sound was heard at the end of this swallow, coinciding with the respiratory expiration that normally occurs. A second patient in this study with a retrospective diagnosis of globus hystericus habitually made a ‘throat clearing’ sound (swallow 10). So two, though not the other 11, non-aspirating swallows showed voicing or harsh glottal release sounds where the expiration would be anticipated. However, removing these swallows from analysis did not markedly improve reliability overall.

Discussion

This is a small study in terms of the number of swallows evaluated. However we believe that there are potential implications for speech and language therapists using this technique at the bedside.

This study has found **cervical auscultation alone to have only fair reliability**, which may be inadequate clinically. Although this study confirms that speech and language therapists have a high true positive detection of aspiration they

Clinical messages

- Speech and language therapists using cervical auscultation alone to detect presence or absence of aspiration in patients with swallowing difficulties show only fair agreement.
- Although raters have a high true positive detection of aspiration, they over detect aspiration when it is not present.
- Increased understanding of the interpretation of the sounds may improve reliability.

Table 1 Cervical auscultation detection of aspiration compared with videofluoroscopy

	Rater 1		Rater 2		Rater 3		Rater 4		Rater 5	
	1	2	1	2	1	2	1	2	1	2
Swallow number										
1 aspiration on VF	1	1	1	1	0	1	1	1	1	1
2 aspiration on VF	1	1	1	1	1	1	1	1	1	1
3	0	0	1	1	1	0	0	0	0	0
4 aspiration on VF	1	1	1	1	0	1	1	1	1	1
5	0	1	1	1	1	1	1	1	1	1
6	1	0	1	1	1	1	1	1	1	1
7	1	1	1	1	1	1	1	1	1	1
8	1	1	1	1	1	0	0	0	1	0
9	0	1	1	1	1	1	1	0	0	0
10	1	0	0	0	1	1	0	0	0	0
11	0	0	1	1	0	1	0	0	0	0
12	1	0	1	1	1	1	1	1	1	0
13	1	1	0	0	1	1	0	0	0	0
14	0	0	0	1	1	1	0	0	1	1
15	0	0	1	1	1	1	0	0	1	1
16	0	0	0	0	0	0	0	0	0	0
Aspirators correct <i>n</i> (%)	3(100)	3(100)	3(100)	3(100)	1(33)	3(100)	3(100)	3(100)	3(100)	3(100)
Non-aspirators correct <i>n</i> (%)	6(46)	5(38)	9(69)	10(77)	10(77)	12(92)	5(38)	4(31)	7(54)	5(38)
Overall correct <i>n</i> (%)	9(56)	8(50)	12(75)	13(82)	11(68)	15(94)	8(50)	7(44)	10(62)	8(50)
Kappa (intra-rater)	0.31		0.85		0.33		0.53		0.72	
Kappa (inter-rater)	0.28									

1 = cervical auscultation interpretation correct compared with videofluoroscopy.

0 = cervical auscultation interpretation incorrect compared with videofluoroscopy.

also **falsely detect aspiration** when it is not present, giving inadequate reliability to the whole test. A speech and language therapist may be fairly sure of accuracy if no aspiration is detected (because there were few false negatives; 2 out of 30 swallow tests (6%) and a high negative predictive value (94%)) but less sure that aspiration is truly present given a positive result, due to the high false positive rate. The **main benefit of cervical auscultation may be accurately ruling out aspiration**, so those patients with a negative result do not require further tests for aspiration although this would need further confirmation in a later study. The findings imply that some patients without true aspiration may be incorrectly classified as aspirating with the result that inappropriate dietary or other intervention measures may be taken. This study calls into question the potential utility of using the sounds of swallowing in isolation as an objective tool, though the effect of cervical auscultation as part of a clinical battery may be different.

It is hypothetically possible that videofluoroscopy misses some aspirations by sight that are detected by sound, resulting in lower accuracy of cervical auscultation for non-aspirating swallows. If this were the case however, the reliability between raters would be higher, because they would agree when detecting those 'invisible aspirations' by sound (assuming the high true positive rate in this study is replicable). In the absence of a better gold standard for aspiration in an individual swallow this cannot currently be further tested.

In this study the prevalence of aspiration, though similar to some dysphagic populations, was fairly low (18%). In patient samples with a higher prevalence rate of aspiration reliability might improve. Reliability is also likely to improve in a mixed population of normal and dysphagic patients but reliability of a screening test should be established in the population of proposed use. It is possible to calculate reliability at various prevalence rates, but this would be unsafe based on the small numbers in this study.

The widely varying intra-scorer reliability results indicated that some therapists were more consistent than others at auscultation. This may be a factor of experience. Therapists with low kappa scores may have been 'guessing' although

their intra-rater scores were better than by chance alone. **One** therapist had 'almost perfect' intra-rater reliability which survives Bonferroni's adjustment for chance ($p = 0.01$) and so is highly unlikely to have been achieved by guessing. This implies she **used useful internal criteria to detect aspiration and non-aspiration by sound**.

The source of error in cervical auscultation is probably multifactorial. Evidence from this study point to both **flawed decision-making** and an unknown factor in some **non-aspirating sounds**. The newly identified previously unsuspected, unreported 'other sounds' at the end of some non-aspirating swallows appear benign (do not mean aspiration has occurred, assuming videofluoroscopy is 100% accurate for detection of aspiration). They may cause error either by **masking clear expiratory breath sounds** that therapists **interpret as indicating no aspiration** or by mimicking the **noisy expiratory breath sounds** that have been found to occur significantly in aspirating swallows.¹⁶ These extra sounds occurred in both patients who are and are not at clinical risk of aspiration. Globus hystericus is a condition not generally regarded as creating a risk of aspiration, and the extra sound may relate to habitual throat clearing. The patient with pseudobulbar palsy is at risk of aspiration but did not actually do so in swallow 16. In this case the extra sound may relate to a learned, compensatory clearing mechanism to quickly remove any potential aspirate. It is possible that speech and language therapists used different internal decision-making criteria to interpret these sounds, with a guessed clinical risk scenario in mind. A speech and language therapist may prefer to risk overdetecting aspiration when it is not present rather than risk missing a positive aspiration, and this may also depend on the perceived clinical scenario. Such possibilities require further investigation.

It is suggested that **cervical auscultation should not be developed as a stand-alone tool** to detect aspiration unless the probability of false positives is understood or the interpretation of sounds can be refined. Further investigations are warranted to establish expert consensus on possible new criteria to interpret the sounds of swallowing, particularly newly described 'benign' sounds.

Acknowledgements

Alison Stroud is supported by a Wales Office of Research and Development studentship. We are thankful to Matthew Cass, Media Resources, UWCM; The Speech and Language Therapy Department at University Hospital of Wales; Dr Robert Newcombe, Department of Epidemiology, Statistics and Public Health, UWCM and Carl Leeke, 3M Healthcare.

References

- 1 Wiles CM. Neurogenic dysphagia. *J Neurol Neurosurg Psychiatry* 1991; **54**: 1037–39.
- 2 Gordon C, Langton Hewer R, Wade DT. Dysphagia in acute stroke. *BMJ* 1987; **295**: 411–14.
- 3 Langmore SE, Terpenning MS, Schork A *et al.* Predictors of aspiration: how important is dysphagia? *Dysphagia* 1998; **13**: 69–81.
- 4 McCullough GH, Rosenbek JC, Robbins JA, Coyle JL, Wood JL. Ordinality and intervality of a penetration-aspiration scale. *J Med Speech-Pathol* 1998; **6**: 65–72.
- 5 Jones B. The pharynx: disorders of function. *Radiol Clin North Am* 1994; **32**: 1103–15.
- 6 Scott A, Perry A, Bench J. A study of interrater reliability when using videofluoroscopy as an assessment of swallowing. *Dysphagia* 1998; **13**: 223–27.
- 7 Selley WG, Flack F, Ellis RE, Brooks WA. Respiratory patterns associated with swallowing: Part 1 The normal adult pattern and changes with age. *Age Ageing* 1989; **18**: 168–72.
- 8 Zenner PM, Losinski DS, Mills RH. Using cervical auscultation in the clinical dysphagia examination in long-term care. *Dysphagia* 1995; **10**: 27–31.
- 9 Eicher PS, Manno CJ, Fox CA, Kerwin ME. Impact of cervical auscultation on accuracy of clinical evaluation in predicting penetration/aspiration in a paediatric population. In: Groher M ed. *Second workshop on cervical auscultation*. Tyson's Corner, McLean, Virginia: Michael Groher, 1994: 28–32.
- 10 McCullough GH, Rosenbek JC, Robbins JA, Coyle JL, Wood JL. Ordinality and intervality of a penetration-aspiration scale. *J Med Speech Pathol* 1998; **6**: 65–72.
- 11 Fleiss JL. *Statistical methods for rates and proportions*, second edition. New York: John Wiley, 1981.
- 12 Landis JR, Koch GG. The measurement of observer agreement for categorical data. *Biometrics* 1977; **33**: 159–74.
- 13 Teasell RW, McRae M, Marchuk Y, Finestone HM. Pneumonia associated with aspiration following stroke. *Arch Phys Med Rehabil* 1996; **77**: 707–709.
- 14 Groher M. The role of the speech-language pathologist in the evaluation and care of oral and pharyngeal dysphagia. *J Neuro Rehabil* 1992; **3**: 5–8.
- 15 Hamlet S, Penney DG, Formolo J. Stethoscope acoustics and cervical auscultation of swallowing. *Dysphagia* 1994; **9**: 63–68.
- 16 Takahashi K, Uyama R, Groher M, Michi K. A new method for detecting dysphagic swallows using acoustic characteristics of soft expiratory sounds. *Dysphagia* 1997; **12**: 115.