International Journal of Speech-Language Pathology, 2012; Early Online: 1–7

informa
healthcare

Intra- and inter-rater reliability for judgement of cough following citric acid inhalation

ANNA MILES[1,2] & MAGGIE-LEE HUCKABEE[1,3]

[1]*University of Canterbury, Christchurch, New Zealand,* [2] *The University of Auckland, Auckland, New Zealand, and* [3]*New Zealand Brain Research Institute, Christchurch, New Zealand*

Abstract
This study investigated the inter-rater and intra-rater reliability of subjective judgements of cough in patients following inhalation of citric acid. Eleven speech-language pathologists (SLPs) currently using cough reflex testing in their clinical practice (experienced raters) and 34 SLPs with no experience using cough reflex testing (inexperienced raters) were recruited to the study. Participants provided a rating of strong, weak, or absent to 10 video segments of cough responses elicited by inhalation of nebulized citric acid. The same video segments presented in a different sequence were re-evaluated by the same clinicians following a 15-minute break. Inter-rater reliability for experienced raters was calculated with a Fleiss' generalized kappa of .487; intra-rater reliability was higher with a kappa of .700. Inexperienced raters showed similar reliability, with kappa values for inter-rater and intra-rater reliability of .363 and .618, respectively. In conclusion, SLPs demonstrate only fair-to-moderate reliability in subjectively judging a patient's cough response to citric acid. Experience in making cough judgements does not improve inter-rater reliability significantly. Further validity and reliability research, including an evaluation of the effect of training on judgement reliability, would be beneficial for guiding clinical policies.

Keywords: *Cough judgement, reliability, cough reflex testing, speech-language pathologist.*

Introduction

Decreased cough sensitivity has been found in neurological diseases such as stroke (Addington, Stephens, Widdicombe, & Rekab, 2005) and bronchopulmonary diseases such as recurrent pneumonia (Niimi, Matsumoto, Ueda, Takemura, Suzuki, & Tanaka, 2003). Diminished voluntary cough is also prevalent in many neurological disorders, including multiple sclerosis (Aiello, Rampello, Granella, Maestrelli, Tzani, Immovilli, et al., 2008), Parkinson's disease (Ebihara, Saito, Kanda, Nakajoh, Takahashi, Arai, et al., 2003; Fontana, Pantaleo, Benvenuti, & Gangemi, 1998; Pitts, Bolser, Rosenbek, Trache, & Sapienza, 2008; Pitts, Troche, Mann, Rosenbek, Okun, & Sapienza, 2010), and duchenne muscular dystrophy (Bach, Ishikawa, & Kim, 1997). It is particularly prevalent in stroke (Addington et al., 2005; Smith-Hammond, Goldstein, Zajac, Gray, Davenport, & Bolser, 2001; Stephens, Addington, & Widdicombe, 2003), with up to 78% of patients with acute unilateral middle cerebral artery infarcts presenting with an abnormal voluntary cough (Stephens et al., 2003). In progressive neurological diseases, cough strength reduces with disease progression (Ebihara et al., 2003). Fifty percent of patients with multiple sclerosis (Aiello et al., 2008) and 40% of those with Parkinson's disease (Nakashima, Maeda, Tabata, Adachi, Kusumi, & Ohshiro, 1997) die from pulmonary complications, with impaired cough considered to be a contributing factor.

Decreased cough reflex sensitivity (Addington et al., 2005; Nakajoh, Nakagawa, Sekizawa, Matsui, Arai, & Sasaki, 2000; Niimi et al., 2003; Sekizawa, Ujiie, Itabashi, Sasaki, & Takishima, 1990) and abnormal voluntary cough (Gauld, 2009; Smith-Hammond, Goldstein, Horner, Ying, Gray, Gonzalez-Rothi, et al., 2009) have both been linked with the development of aspiration pneumonia. Clinically, it is important to distinguish between cough sensitivity and cough strength as they are different neurophysiological processes and, although both may lead to an increased risk of developing pneumonia, assessment and management may be significantly different.

Cough sensitivity is the degree to which sensory receptors in the larynx and tracheobronchial tree detect irritation. Vagal sensory afferent nerves are activated, sending an excitatory signal to neurones in the nucleus tractus solitarius (NTS) (Fontana & Lavorini, 2006). Decreased cough sensitivity is likely to lead to silent aspiration of food/fluids and failed initiation of airway protection.

Correspondence: Anna Miles, Speech Sciences, Tamaki Innovation Campus, The University of Auckland, Private Bag 92019, Auckland 1142, New Zealand. Tel: 006495222620. Fx: 006493737043. Email: a.miles@auckland.ac.nz

ISSN 1754-9507 print/ISSN 1754-9515 online
Published by Informa UK, Ltd.
DOI: 10.3109/17549507.2012.692812

Cough strength is the amplitude of the motor pattern of coughing either in response to the stimulation or under voluntary control. To generate an effective cough, muscles of both inspiration and expiration are needed (Gauld, 2009) involving the intrinsic laryngeal muscles, diaphragm, intercostal and ribcage muscles and abdominal muscles (Fontana & Lavorini, 2006). Decreased cough strength is likely to lead to a consequent inability to clear material from the airway and is therefore likely to exacerbate pulmonary consequences resulting from penetration/aspiration (Fontana & Widdicombe, 2007). Both cough sensitivity and cough strength are therefore important to the assessment of dysphagia and an assessment protocol which evaluates both may be advantageous (Widdicombe, Addington, Fontana, & Stephens, 2011).

Many researchers have investigated the relationship between objective measures of cough strength and aspiration risk. Pitts et al. (2010) found that objective airflow measures from a voluntary cough may identify patients who are at-risk of penetration and aspiration in Parkinson's disease. Smith-Hammond et al. (2001, 2009) have published various studies correlating abnormal voluntary cough with aspiration risk in dysphagia post-stroke. They have discovered that objective measures of cough, such as peak flow of inspiratory phase and expulsive phase and cough volume acceleration, are impaired in patients who aspirate.

Interestingly, other researchers have warned against the use of these measures of voluntary cough strength for judging a patient's ability to clear the airway after penetration/aspiration of food/fluids (Addington, Stephens, Phelipa, Widdicombe, & Ockey, 2008). Voluntary and reflexive coughs respond differently to disease (Fontana et al., 1998; Stephens et al., 2003; Ward, Seymour, Jolley, Polkey, Kaira, & Moxham, 2010). The motor component of voluntary cough has been proven different to that of a reflexive cough with a suggestion of different underlying neural pathways (Lasserson, Mills, Arunachalam, Polkey, Moxham, & Kalra, 2006; Magni, Chellini, Lavorini, Fontana, & Widdicombe, 2011). A voluntary cough is cortically-mediated and is used to clear the airway prior to speech or for clearing the airways once material is present in the tracheobronchial tree, whereas an expiratory reflex followed by a reflexive cough is a brainstem-mediated response to airway irritation or threat, i.e., penetration/aspiration/obstruction (Addington et al., 2008). Measures of reflexive cough efficiency may be more useful for identifying at-risk patients for aspiration pneumonia in view of its importance as the initial airway protection mechanism (Magni et al., 2011).

There are differences of opinion among research groups on the adequacy of subjective cough measures, perhaps related to their primary research focuses. Addington and Widdecombe (2009), who have focused their work on the sensory cough pathway, write "a subjective measurement of voluntary cough (VC) has not been shown to give less acceptable results than the objective method" (Addington & Widdecombe, 2009, p. 647), but provide no empirical justification for this statement. In contrast, Gauld (2009) and Smith-Hammond and Goldstein (2006), whose work has focused on the motor pathway/cough strength, write of their opinion that subjective assessment of cough is not accurate or reliable, but again do not support this statement with data.

Objective cough strength measures are not a standard component of a bedside swallowing assessment, whereas subjective cough strength judgements are common in clinical practice. However, there is a paucity of research looking specifically at the inter-rater reliability of subjective judgements of either voluntary or reflexive cough. In a literature search of cough judgement research, three research groups were found to have assessed inter-rater reliability. These studies all investigated clinical swallowing evaluation protocols with voluntary cough as one of many measures. Daniels, Brailey, Priestly, Herrington, Weiberg, and Foundas (1998) evaluated the reliability of clinical swallowing assessment measures, including normal vs abnormal voluntary cough and reported 95% inter-rater agreement for all binary measures. In a subsequent publication, they described their definition of abnormal voluntary cough as "a weak response, verbalized response, or no response on [when] given the command to cough" (Daniels, Lindsay, Ballo, Mahoney, & Foundas, 2000, p. 1031). Rater experience or training was not discussed. Rosenbek, McCullough, and Wertz (2004) also found high inter-rater reliability between three experienced judges across all aspects of a clinical swallowing evaluation, including voluntary cough strength and quality (wet vs dry). Again they give few details on the training of the clinicians on these judgements. McCullough, Rosenbek, Wertz, and McCoy (2005), again, used only experienced judges. They judged volitional cough strength with 100% agreement, volitional cough quality with 85% agreement, reflexive cough strength with 85% agreement, and reflexive cough quality with 92% agreement.

No other research has been identified which evaluates the reliability of reflexive cough judgement. Therefore, this small pilot study investigated the inter- and intra-rater reliability of SLPs, and the influence of practice, in subjective judgement of involuntary cough during a cough reflex test.

Method

Participants

The reliability study, including use of video recordings of patients who provided informed consent, was reviewed and approved by an appropriate regional ethics committee. Participants included 11 SLPs who received an 8-hour cough reflex testing training session and had been using cough reflex testing for

1 year (experienced raters), and 34 SLPs with no experience or formal training with the cough reflex test (inexperienced raters). Recruitment, consent, and participation took place at a number of professional development events. The SLPs ranged in clinical experience from new graduate to specialist, and all reported working with adults with dysphagia.

Materials

Informed consent for videotaping was secured from 10 hospitalized adults (mixed gender, age range 24–91 years) at two urban hospitals to provide the data for reliability analysis. The videos represented 10 consecutive dysphagia swallowing assessments. Patient aetiology included progressive neurological disease, acute stroke, and frail elderly. The strength of patients' coughs was not objectively assessed, but the 10 patients were considered by consensus of the authors to be representative of the range of responses seen in a hospital setting. Non-dysphagic patients were not selected so that the videos represented an "average" caseload. Patients were video-recorded using a Mino HD flip video camera (CISCO, Irvine, CA) while undergoing a cough reflex test as part of their standard dysphagia assessment. The 10 video clips were edited into a 3-minute high-resolution movie using iMovie (Apple, Cupertino, CA). Each clip showed a patient receiving one 15-second dose of nebulized citric acid solution (diluted in 0.9% sodium chloride) through a facemask using a Pulmo-Mate Compressor/Nebulizer (model 4650I) (DeVilbiss Healthcare LLC, PA). This movie was labelled First Viewing. The movie was then recreated, showing the same 10 video clips in a randomly different order and labelled Second Viewing.

Procedure

Participants viewed the movies through Windows Media Player via projector onto a conference room screen. Participants were asked to independently rate the cough response seen in each video clip as strong (two or more strong coughs), weak (two or more weak coughs), or absent (one or no cough). The C2 scoring system (Morice et al., 2007) was chosen for this study. This system requires a response of two coughs within 15 seconds of presentation of tussive stimuli, and is recommended by the European Respiratory Society (ERS) Guideline for Assessment of Cough (Morice et al., 2007). Participants were given no specific definition of a cough, but were told that a throat clear is not considered a cough response. They were provided with no further education about cough or cough strength judgement. After watching the "First Viewing" of the movie they were given a 15-minute break where participants were discouraged from discussing the videos. They were then asked to watch the "Second Viewing" of the movie. After the second viewing, participants were encouraged to discuss their experience of judging the videos with each other and the researchers.

Data analysis

Data were analysed using SPSS version 20 software (SPSS, Chicago, IL). Fleiss' generalized kappa measurement of agreement for multiple raters when assigning categorical ratings was used. The data from experienced raters and inexperienced raters were analysed as separate groups. Numerical coding for the three categories was defined as: strong (2), weak (1), and no cough (0). Additionally, within each group, each category was separately analysed for agreement. The Landis and Koch (1977) definition of levels of Fleiss' generalized kappa agreement was used for interpretation. Descriptive data from the unstructured discussion between participants following completion of the viewings were collected. Using the theory of qualitative analysis, the individual comments were divided into meaning units, and themes were derived (Graneheim & Lundman, 2004).

Results

The raw agreement data for all raters for first viewing and second viewing are presented in Tables I and II. The inexperienced raters displayed an overall

Table I. Raw agreement data (first viewing).

	Experienced raters				Inexperienced raters			
	absent	weak	strong	% agreement	absent	weak	strong	% agreement
Video 1	0	1	10	90	0	12	22	65
Video 2	0	10	1	90	2	16	16	47
Video 3	0	7	4	63	0	12	22	65
Video 4	8	2	1	73	20	14	0	59
Video 5	2	5	4	45	4	8	22	65
Video 6	0	6	5	55	0	12	22	65
Video 7	0	2	9	82	1	11	22	65
Video 8	7	4	0	64	20	14	0	65
Video 9	10	1	0	91	31	2	1	92
Video 10	11	0	0	100	33	1	0	97
Mean agreement				75				69

Table II. Raw agreement data (second viewing).

	Experienced raters				Inexperienced raters			
	absent	weak	strong	% agreement	absent	weak	strong	% agreement
Video 1	0	0	11	100	0	5	29	85
Video 2	0	7	4	64	0	18	16	53
Video 3	0	7	4	64	1	16	17	50
Video 4	7	4	0	64	28	6	0	82
Video 5	1	4	6	55	4	7	23	68
Video 6	0	4	7	64	0	16	18	53
Video 7	0	3	8	73	0	9	25	73
Video 8	4	7	0	64	22	12	0	65
Video 9	11	0	0	100	31	3	0	92
Video 10	11	0	0	100	33	1	0	97
Mean agreement				75				72

agreement of 69% (range 47–97%) for the first viewing and 72% (range 50–97%) for the second viewing. In comparison, the experienced raters displayed an overall agreement of 75% for both viewings (first viewing range 45–100%, second viewing range 55–100%), with three videos rated with 100% agreement in the second viewing.

Inter-rater reliability

Inter-rater reliability for inexperienced raters for the first viewing was calculated at $\kappa = .363$ ($p < .001$; 95% CI = .345–.382). For the second viewing, these raters achieved only a marginally greater $\kappa = .407$ ($p < .001$; 95% CI = .389–.426). The experienced raters achieved a $\kappa = .459$ ($p < .001$; 95% CI = .400–.518) across the first viewing. This was marginally greater for the second viewing of the videos with a $\kappa = .488$ ($p < .001$; 95% CI = .428–.546). There were marked differences in agreement between categories for both groups on both viewings. A summary of overall agreement and category specific agreement is presented in Table III, including κ-values and 95% confidence intervals. Across both groups, agreement for strong cough ranged from $\kappa = .375$–.488, weak cough agreement ranged from $\kappa = .067$–.293, and absent cough agreement ranged from $\kappa = .634$–.699. Confidence intervals for category agreement were wide, likely secondary to the small data set once the 10 items were analysed separately as three categories.

Intra-rater reliability

Intra-rater reliability was higher than inter-rater agreement with a $\kappa = .618$ ($p < .001$; 95% CI = .543–.694) for inexperienced raters and a $\kappa = .700$ ($p < .001$; 95% CI = .568–.813) for experienced raters. Again, differences in agreement across categories were seen. Inexperienced raters intra-rater agreement ranged from; strong $\kappa = .662$ (95% CI = .370–.954), weak $\kappa = .409$ (95% CI = .126–.691), no cough $\kappa = .763$ (95% CI = .477–1.00). Experienced raters intra-rater agreement ranged from; strong $\kappa = .707$ (95% CI = .206–1.00), weak $\kappa = .580$ (95% CI = .074–1.00), no cough $\kappa = .816$ (95% CI = .313–1.00).

Qualitative responses

A list of qualitative responses is presented in Table IV. The raters' responses were separated into three

Table III. Overall agreement and category-specific agreement.

		κ	*p*-value	Lower 95% CI	Upper 95% CI
Experienced raters, viewing 1	Overall agreement	.459	.00000	.400	.518
	Strong cough rating	.430	.003	.127	.732
	Weak cough rating	.268	.05	−.054	.590
	Absent cough rating	.678	.00002	.356	1.00
Experienced raters, viewing 2	Overall agreement	.487	.00000	.43	.55
	Strong cough rating	.488	.001	.176	.800
	Weak cough rating	.293	.04	−.034	.620
	Absent cough rating	.690	.00001	.383	1.0
Inexperienced raters, viewing 1	Overall agreement	.363	.00000	.345	.382
	Strong cough agreement	.375	.00001	.202	.548
	Weak cough agreement	.067	.19	−.081	.216
	Absent cough agreement	.634	.00000	.477	.791
Inexperienced rater, viewing 2	Overall agreement	.407	.00000	.389	.426
	Strong cough agreement	.398	.00000	.225	.571
	Weak cough agreement	.083	.12	-.057	.223
	Absent cough agreement	.699	.00000	.533	.865

Table IV. Themes of qualitative data; representative examples of comments from raters.

Definition of a cough	Definition of a weak cough	Practice/confidence
"I decided to close my eyes and just listen for the sound of a cough as the picture confused me"	"That was much harder than I thought it would be, I think I judge it on the person's general strength clinically, not their cough"	"It felt a lot easier the second time around"
"He looked like he was coughing but I don't know"	"What is a weak cough … all my patients are weak?"	"I remember when I first started using the cough reflex test, I was confused by weak vs strong but now that it is something I focus on more, I feel more confident"
"What is a cough?"	"I don't know … he didn't cough as strong as I can but it seemed strong enough for his age"	"I don't think I thought about cough strength before I started using the cough reflex test"
"His shoulders moved but I didn't know if was a cough or not … I thought I'd just need to count the coughs but it was more complicated than that!"	"He looked so weak … it had to be a weak"	
"What is the difference between a throat clear and a cough?"		

themes: (i) definition of cough; (ii) definition of weak cough; (iii) practice/confidence. Participants discussed uncertainty in the definition of cough: "he looked like he was coughing but I don't know", the definition of a weak cough "he looks very frail, does that make him weak?", and voiced increase in confidence between viewings "I had a better idea about what I was going to rate as a cough the second time".

Discussion

Inter/intra-rater reliability

SLPs in this pilot study achieved fair-to-moderate agreement (κ-values ranging between .363–.487 when making subjective judgements of reflexive cough without training (Landis & Koch, 1977). This agreement is comparable to other areas of SLP where subjective judgements are used. Inter/intra-rater reliability in videofluoroscopic studies of swallowing (VFSS) has been researched extensively. Rosenbek, Robbins, Roecker, Coyle, and Woods (1996) found high agreement for judges using the Penetration-Aspiration Scale [ICC 0.96], but many other researchers have found lower levels of agreement. Ekberg, Nylander, Fork, Sjoberg, Birch-Iensen, and Hillarp (1988) found large variation across VFSS measures with a Kappa value of .57 for presence of penetration and Kuhlemeier, Yates, and Palmer (1998) found only 30% agreement for abnormal ratings and 78% agreement for normal ratings of VFSS analysis. Similar variations in agreement levels have been seen in perceptual voice assessment, where agreement can range from high-to-low depending on the measure, the type of rating (i.e., binary vs rating scale), and the training/experience of the rater (Kreiman, Gerratt, Kempster, Erman, & Berke, 1993; Oates, 2009; Webb, Carding, Deary, Mackenzie, Steen, & Wilson, 2004).

Both inexperienced and experienced SLPs appeared more reliable at judging a "cough" vs "no cough", and agreement was lower for judging a "weak" vs "strong"' cough with 100% agreement for videos 9 and 10 (absent cough), whereas only 50–64% agreement between strong and weak cough for videos 2, 3, and 6. Although confidence intervals were wide, when the data were analysed by category, the raters demonstrated greater agreement for absent cough (κ-value ranges across groups and viewings = .634–.699) compared with weak cough (κ-value range across groups and viewings = .375–.488). Raters commented on the difficulty in judging strength during the study. They reported that they did not know the definition of a cough and that they were confused about how to discriminate it from other responses such as a throat clear.

Although experience in cough reflex testing was controlled, the overall clinical experience of the participants was not controlled for or collected. The impact of clinical experience of dysphagia management on intra- and inter-rater reliability cannot therefore be determined. A clinician's own "internal standards" and how this influences clinical judgement has been extensively researched in the area of the perceptual evaluation of voice (Gerratt, Krieman, Antonanzas-Barroso, & Berke, 1993; Oates, 2009). Clinician's experience and training could likewise influence cough intensity judgement. A clinician who works with a neurologically impaired/elderly population is likely to perceive intensity of cough differently to a clinician working in an outpatient ENT clinic.

In keeping with the C2 scoring method, video clips showed only 15 seconds of a patient's response to tussive stimuli. Judgements may be better or more representative of clinical practice if longer observation periods were given. In clinical practice, consistency of response would be determined over three trials before making a final judgement. Alternatively, the patient's physical appearance may adversely affect judgement of strength. One video presented a frail gentleman lying in bed. He was judged by the researchers to have a strong, prompt cough, but raters showed significant variability in their judgements between strong and weak. Another potential barrier to the use of cough reflex testing to judge strength is the noise of the nebulizer affecting the ability to hear the cough clearly.

Practice effect

SLPs may not consistently receive training in cough physiology, and some clinicians appeared not to have a clear understanding of the definition of a cough. Experience and training has shown to improve perceptual judgements in other areas of SLP such as voice (Oates, 2009) and a videofluoroscopic study of swallowing (Logemann, Lazarus, Keeley, Sanchez, & Rademaker, 2000). Further training in what a cough is may increase reliability in judgements. All of the experienced raters had received initial training in the cough reflex test before initiating application of this method in clinical work, but this training focused on the cough reflex and had not included a definition of cough or a definition of a strong vs a weak cough. The descriptive data in this study indicates that both novice and experienced SLPs lack confidence in defining a cough and defining a weak vs strong cough. Participants described feeling more confident when watching the videos for a second time, but agreement remained fair-to-moderate. University and in-hospital training could focus on more consistent definitions on a reflexive cough, including examples of strong, weak, and absent cough responses.

Despite 1 year of cough reflex testing, the inter- and intra-rater reliability of experienced raters was not significantly different to the inexperienced raters overall. There was a slight trend towards more consistent judgements in the experienced group (inexperienced raters $\kappa = 0.407$, experienced raters $\kappa = 0.488$), with higher intra-rater reliability in the experienced group also (inexperienced raters $\kappa = 0.618$, experienced raters $\kappa = 0.700$). This suggests that practice may lead to increased confidence. In allowing participants to talk after the first viewing, bias could have been added but there is clearly little impact of this, as participants did not markedly change their ratings across viewings.

Validity

Objective cough strength measures were not used in this study, so no conclusion can be made about the accuracy of SLPs judgements, only their agreement. A comparison of an objective reflexive cough measure (such as airflow measures) vs subjective judgement of cough strength would add further to this area of research.

Research has shown both abnormal volitional cough strength and reduced cough sensitivity as risk factors for pneumonia (Addington, Stephens, & Gilliland, 1999; Smith-Hammond et al., 2009). If a cough reflex test (pass/fail) with a subjective strength rating (weak/strong) can be shown as reliable, this may prove a beneficial additional tool to the clinical swallowing evaluation.

Clinically, there are difficulties in assessing voluntary cough in patients after stroke secondary to barriers of impaired cognition, communication, and apraxia (Addington & Widdecombe, 2009; Stephens et al., 2003). A judgement of reflexive cough strength may certainly be more useful for identifying those at risk of being unable to clear aspirated material than a voluntary cough judgement (Magni et al., 2011). In view of the high inter-rater reliability found in experienced researchers in judging volitional cough strength (Daniels et al., 1998; Rosenbek et al., 2004), perhaps with more directed training, clinicians could also reach greater agreement in reflexive cough judgement.

Conclusions

This pilot study highlights the need for further controlled studies, but also presents an area of need for professional development for SLPs. Future studies on cough judgement could probe issues of reliability such as what classifies a weak cough using more controlled video materials. We did not address validity of subjective cough judgement in this study, and future studies where participants are rating coughs of a known strength would be valuable. Untrained SLPs were only fair-to-moderately reliable in subjectively judging a patient's cough. Experience in making cough judgements did not improve inter-rater reliability significantly. The effect of years of clinical practice in dysphagia and type of clinical experience are worth further investigation. Studying the effect of training on subjective cough judgement would add to this area of research.

Acknowledgements

The authors appreciate the contributions of all the SLPs who participated in this study and the conference organizers who allowed recruitment to take place at their professional development events.

Declaration of Interest: The authors report no conflicts of interest. The authors alone are responsible for the content and writing of the paper.

References

Addington, W., & Widdecombe, J. (2009). Cough and aspiration: Letter to editor. *Chest*, *136*, 647.

Addington, W., Stephens, R., & Gilliland, K. (1999). Assessing the laryngeal cough reflex and the risk of developing pneumonia after stroke: An interhospital comparison. *Stroke*, *30*, 1203–1207.

Addington, W., Stephens, R., Phelipa, M., Widdicombe, J., & Ockey, R. (2008). Intra-abdominal pressures during voluntary and reflex cough. *Cough*, *4*, 1–9.

Addington, W., Stephens, R., Widdicombe, J., & Rekab, K. (2005). Effect of stroke location on the laryngeal cough reflex and pneumonia risk. *Cough*, *1*, 4–10.

Aiello, M., Rampello, A., Granella, F., Maestrelli, M., Tzani, P., Immovilli, P., et al. (2008). Cough efficacy is related to the disability status in patients with multiple sclerosis. *Respiration*, *76*, 311–316.

Bach, J., Ishikawa, Y., & Kim, H. (1997). Prevention of pulmonary morbidity for patients with duchenne muscular dystrophy. *Chest*, *112*, 1024–1028.

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

Daniels, S., Lindsay, M., Ballo, A., Mahoney, M., & Foundas, A. (2000). Clinical predictors of dysphagia and aspiration risk: Outcome measures in acute stroke patients. *Archives of Physical Medicine and Rehabilitation*, *81*, 1030–1033.

Ebihara, S., Saito, H., Kanda, A., Nakajoh, M., Takahashi, H., Arai, H., et al. (2003). Impaired efficacy of cough in patients with Parkinson disease. *Chest*, *124*, 1009–1015.

Ekberg, O., Nylander, G., Fork, F.-T., Sjoberg, S., Birch-Iensen, M., & Hillarp, B. (1988). Interobserver variability in cineradiographic assessment of pharyngeal function during swallow. *Dysphagia*, *3*, 46–48.

Fontana, G., & Lavorini, F. (2006). Cough motor mechanisms. *Respiratory Physiology and Neurobiology*, *152*, 266–281.

Fontana, G., & Widdicombe, J. (2007). What is cough and what should be measured? *Pulmonary Pharmacology and Therapeutics*, *20*, 307–312.

Fontana, G., Pantaleo, T., Benvenuti, F., & Gangemi, S. (1998). Defective motor control if coughing in Parkinson's disease. *American Journal of Respiratory Critical Care Medicine*, *158*, 458–464.

Gauld, L. (2009). Airway clearance in neuromuscular weakness. *Developmental Medicine and Child Neurology*, *51*, 350–355.

Gerratt, B., Krieman, J., Antonanzas-Barroso, N., & Berke, G. (1993). Comparing internal and external standards in voice quality judgements. *Journal of Speech and Hearing Research*, *36*, 14–20.

Graneheim, U., & Lundman, B. (2004). Qualitative content analysis in nursing research: Concepts, procedures and measures to achieve trustworthiness. *Nurse Education Today*, *24*, 105–112.

Kreiman, J., Gerratt, B., Kempster, G., Erman, A., & Berke, G. (1993). Perceptual evaluation of voice quality: Rreview, tutorial and a framework for future research. *Journal of Speech and Hearing Research*, *36*, 21–40.

Kuhlemeier, K., Yates, P., & Palmer, J. (1998). Intra- and inter-rater variation in the evaluation of videofluoroscopic studies. *Dysphagia*, *13*, 142–147.

Landis, J. R., & Koch, G. (1977). The measurement of observer agreement for categorical data. *Biometrics*, *33*, 159–174.

Lasserson, D., Mills, K., Arunachalam, R., Polkey, M., Moxham, J., & Kalra, L. (2006). Differences in motor activation of voluntary and reflex cough in humans. *Thorax*, *61*, 699–705.

Logemann, J., Lazarus, C., Keeley, S. P., Sanchez, A., & Rademaker, A. (2000). Effectiveness of four hours of education in interpretation of radiographic studies. *Dysphagia*, *15*, 180–183.

Magni, C., Chellini, E., Lavorini, F., Fontana, G., & Widdicombe, J. (2011). Voluntary and reflex cough: Similarities and differences. *Pulmonary Pharmacology and Therapeutics*, *24*, 308–311.

McCullough, G., Rosenbek, J., Wertz, R., & McCoy, C. (2005). Utility of clinical swallowing examination measures for detecting aspiration post-stroke. *Journal of Speech, Language, and Hearing Research. Rockville*, *48*, 1280–1293.

Morice, A., Fontana, G., Belvisi, M., Birring, S., Chung, K., Dicpinigaitis, P., et al. (2007). ERS guidelines on the assessment of cough. *European Respiratory Journal*, *29*, 1256–1276.

Nakajoh, K., Nakagawa, T., Sekizawa, K., Matsui, T., Arai, H., & Sasaki, H. (2000). Relation between incidence of pneumonia and protective reflexes in post-stroke patients with oral or tube feeding. *Journal of Internal Medicine*, *247*, 39–42.

Nakashima, K., Maeda, M., Tabata, M., Adachi, Y., Kusumi, M., & Ohshiro, H. (1997). Prognosis of Parkinson's disease in Japan. *European Journal of Neurology*, *38*(Suppl. 2), 60–63.

Niimi, A., Matsumoto, H., Ueda, T., Takemura, M., Suzuki, K., & Tanaka, E. (2003). Impaired cough reflex in patients with recurrent pneumonia. *Thorax*, *58*, 152–153.

Oates, J. (2009). Auditory-perceptual evaluation of disordered voice quality. Pros, cons and future directions. *Folia Phoniatrica et Logepaedica*, *61*, 49–56.

Pitts, T., Bolser, D., Rosenbek, J., Trache, M., & Sapienza, C. (2008). Voluntary cough production and swallow dysfunction in Parkinson's disease. *Dysphagia*, *23*, 297–301.

Pitts, T., Troche, M., Mann, G., Rosenbek, J., Okun, M., & Sapienza, C. (2010). Using voluntary cough to detect penetration and aspiration during oropharyngeal swallowing in patients with Parkinson disease. *Chest*, *138*, 1426–1431.

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

Rosenbek, J., Robbins, J., Roecker, E., Coyle, J., & Woods, J. (1996). A penetration- aspiration scale. *Dysphagia*, *11*, 93–98.

Sekizawa, K., Ujiie, Y., Itabashi, S., Sasaki, H., & Takishima, T. (1990). Lack of cough reflex in aspiration pneumonia. *The Lancet*, *335*, 1228–1229.

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

Smith-Hammond, C., Goldstein, L., Horner, R., Ying, J., Gray, L., Gonzalez-Rothi, L., et al. (2009). Predicting aspiration in patients with ischemic stroke comparison of clinical signs and aerodynamic measures of voluntary cough. *Chest*, *135*, 769–777.

Smith-Hammond, C., Goldstein, L., Zajac, D., Gray, L., Davenport, P., & Bolser, D. (2001). Assessment of aspiration risk in stroke patients with quantification of voluntary cough. *Neurology*, *56*, 502–506.

Stephens, R., Addington, W., & Widdicombe, J. (2003). Effect of acute unilateral middle cerebral artery infarcts on voluntary cough and the laryngeal cough reflex. *American Journal of Physical Medicine and Rehabilitation*, *May*, 379–393.

Ward, K., Seymour, J., Jolley, C., Polkey, M., Kaira, L., & Moxham, J. (2010). Acute ischaemic hemispheric stroke is associated with impairment of reflex in addition to voluntary cough. *European Respiratory Journal*, *36*, 1383–1390.

Webb, A., Carding, P., Deary, I., Mackenzie, K., Steen, N., & Wilson, J. (2004). The reliability of three perceptual evaluation scales of dysphonia. *European Archives of otorhinolaryngology*, *216*, 429–434.

Widdicombe, J., Addington, W., Fontana, G., & Stephens, R. (2011). Voluntary and reflex cough and the expiration reflex; implications for aspiration after stroke. *Pulmonary Pharmacology and Therapeutics*, *24*, 312–317.