International Journal of Language & Communication Disorders

INT J LANG COMMUN DISORD, JANUARY/FEBRUARY 2021,
VOL. 56, NO. 1, 174–204

Review

Utility of ultrasound in the assessment of swallowing and laryngeal function: A rapid review and critical appraisal of the literature

Jodi E. Allen†, Gemma M. Clunie‡§, Claire Slinger¶, Jemma Haines‖#, Corinne Mossey-Gaston**, Chariss J. Zaga††, Becky Scott**, Sarah Wallace# and Roganie Govender‡‡SSSS

†National Hospital for Neurology & Neurosurgery, University College London Hospitals NHS Foundation Trust, London, UK
‡National Centre for Airway Reconstruction, Department of Otolaryngology, Head and Neck Surgery, Imperial College Healthcare NHS Trust, London, UK
§Department of Surgery and Cancer, Imperial College London
¶Lancashire Teaching Hospitals, Preston, Lancashire, UK
‖NIHR Biomedical Research Centre, The University of Manchester, Manchester, UK
#Wythenshawe Hospital Manchester University NHS Foundation Trust, Manchester, UK
**Royal Papworth Hospital, Cambridge, UK
††Austin Health, Heidelberg, Australia
‡‡University College London Hospitals NHS Foundation Trust, London, UK
*SSSS*Head & Neck Academic Centre, University College London, London, UK

(*Received August 2020; accepted October 2020*)

Abstract

Background: Ultrasound (US) is not widely used as part of the speech and language therapy (SLT) clinical toolkit. The COVID-19 pandemic has intensified interest in US as an alternative to SLT instrumental tools such as the videofluoroscopic swallowing study (VFSS), fibreoptic endoscopic evaluation of swallowing (FEES) and endoscopic evaluation of the larynx (EEL) as a non-invasive, non-aerosol-generating procedure that can be delivered at the bedside to assess swallowing and/or laryngeal function. To establish the appropriacy of routine US use, and in response to a national professional body request for a position statement, a group of expert SLTs conducted a rapid review of the literature.
Aim: To explore critically the clinical utility of US as an assessment tool for swallowing and laryngeal function in adults.
Methods & Procedures: A rapid review of four databases was completed to identify articles using US to assess swallowing and/or laryngeal function in adults compared with reference tests (VFSS/FEES/EEL/validated outcome measure). Screening was completed according to predefined inclusion/exclusion criteria and 10% of abstracts were rescreened to assess reliability. Data were extracted from full texts using a predeveloped form. The QUADAS-2 tool was used for quality ratings. Information from included studies was summarized using narrative synthesis and visual illustration.
Outcomes & Results: Ten papers used US to assess swallowing, and 13 to assess laryngeal function. All were peer-reviewed primary studies across a range of clinical populations and with a wide geographical spread. Four papers had an overall low risk of bias, but the remaining 19 had at least one domain where risk of bias was judged as high or unclear. Applicability concerns were identified in all papers. The papers that used US to assess swallowing varied widely in terms of the anatomical structures assessed and methodology employed. The papers assessing laryngeal function were more homogenous in their methodology. Sensitivity and specificity data were provided for 12 of the laryngeal function papers with ranges of 64.3–100% and 48.5–100%, respectively.
Conclusions & Implications: There is burgeoning evidence to support the use of US as an adjunct to SLT clinical assessment of swallowing and laryngeal function. However, the current literature does not support its use as a

Address correspondence to: Jodi Allen, University College London Hospitals NHS Foundation Trust, National Hospital for Neurology & Neurosurgery, Therapy and Rehabilitation, 8–11 Queen Square, London WC1N 3BG, UK; email: Jodi.allen@nhs.net

International Journal of Language & Communication Disorders
ISSN 1368-2822 print/ISSN 1460-6984 online
DOI: 10.1111/1460-6984.12584

tool in isolation. Further research is required to establish reliability in US assessment as well as clear SLT-driven protocols and training.

Keywords: acquired, adults, dysarthria, dysphagia, neurodegenerative diseases.

What this paper adds

What is already known on the subject

- US has demonstrated potential as an assessment tool for objective parameters of swallowing. Its use for laryngeal assessment (gross vocal fold movement) is also widely recognized within the literature. This review appraised the literature related to US as an alternative or adjunctive tool for the assessment of swallowing and laryngeal function.

What this paper adds to existing knowledge

- This paper identifies that the current evidence base for US as a swallowing or laryngeal assessment tool is heterogenous and of variable quality. No study combined the assessment of swallowing and laryngeal function, and only two studies assessed more than one parameter of swallowing, limiting the clinical application of the results.

What are the potential or actual clinical implications of this work?

- This review shows that US is a non-invasive accessible tool that can offer a detailed focal assessment of swallowing and laryngeal function, such as hyoid displacement and vocal fold mobility. With the development of protocols, training packages and competency standards, US has the potential to be used as an adjunct to SLT assessment of swallowing and laryngeal function. There is not currently enough evidence to support the use of US as a stand-alone tool for SLT assessment of swallowing or laryngeal function.

Introduction

Difficulties with swallowing (dysphagia) and laryngeal function comprise a large proportion of the caseloads of speech and language therapists (SLTs). The assessment of laryngeal function is an essential component of the swallowing assessment because of its role in airway protection and cough (Pitts 2014). This is particularly true in populations where the underlying disease has multi-system effects, for example, patients with respiratory, neurological or neuromuscular conditions (Pitts *et al*. 2008, Bourke 2014, McGrath *et al*. 2020).

The clinical management of dysphagia and laryngeal impairment relies on thorough information-gathering. This includes a detailed case history, direct examination, perceptual evaluation and diagnostic tests (Suiter and Gosa 2019). SLTs use instrumental assessments to gain objective information about the functional anatomy of key structures and their related biomechanics. They are an essential part of the SLT toolkit to guide diagnostics, evidence-based decision-making, goal-setting and rehabilitation. The most routinely used instrumental assessments include videofluoroscopic swallowing study (VFSS), flexible endoscopic evaluation of swallowing (FEES) and endoscopic evaluation of the larynx (EEL) (Martin-Harris and Jones 2008, Wallace *et al*. 2020, Jones *et al*. 2020). While these tools offer clear imaging of swallowing and laryngeal biomechanics, measurement of movement is cumbersome and requires image extraction to external software to improve reliability. The invasive nature of FEES and EEL limits accessibility and VFSS must be conducted in an upright posture in a radiology suite.

The COVID-19 pandemic has restricted access and provision of standard SLT procedures (VFSS, FEES, EEL) due to the risk of increased aerosol generation and disease transmission (Tran *et al*. 2012, Bolton *et al*. 2020). SLTs are therefore exploring alternative lower risk tools to support the assessment of swallowing and laryngeal biomechanics.

Use of ultrasound for the assessment of swallowing and laryngeal function

An ultrasound (US) scan is a procedure that uses high-frequency sound waves to capture images by placing a sound-emitting transducer directly onto the skin. This collects echoes reflected by the body part and transforms them into decoded signals to form an image (Aldrich 2007). US has been used to study tongue, hyoid and laryngeal movement in swallowing (Shawker *et al*. 1984, Chi-Fishman 2005, Nakamori *et al*. 2016), laryngeal function in post-surgical populations (da

Costa *et al.* 2019) and guide extubation of patients in critical care (Ruan *et al.* 2018). It has not, however, been adopted into routine SLT clinical practice.

A Brazilian review (Leite *et al.* 2014) identified published studies using US to assess swallowing in adults and paediatrics between August 2002 and 2013. The review summarized 17 studies, of which 12 were based on an adult population. Hyoid bone movement was the most explored swallowing parameter, but methodological variability prevented any firm conclusions. Many studies used US as an outcome measure to assess differences between groups of different age or condition. Less than one-quarter of included studies validated US against reference tools such as VFSS, FEES or EEL, limiting applicability to SLT. The authors reported that US was a fast, non-invasive, low-cost method for evaluating objective parameters of swallowing but made no recommendations for the use of US within SLT practice. Since this review there has been a considerable advancement in US technology and interest in its clinical application, warranting an updated review.

US assessment of laryngeal function has also been described fairly extensively within the literature (Noel *et al.* 2020) and reported to be a viable method to assess vocal fold function in a post-thyroidectomy population (Da Costa *et al.* 2019). Previous reviews made recommendations for further research into the use of US for the assessment of swallowing and laryngeal function but without guidelines for implementation within clinical practice. The speed and portability, as well as overall safety and lack of radiation requirement, support the potential for wide application of US, however limited evidence, and no obvious investment in training and skill acquisition, means US has not gained the same prominence as other tools such as VFSS and FEES.

The primary aim of this study was to explore the clinical utility of US as an assessment tool compared with gold standard routine SLT assessment tools in adults both with and without suspected swallowing or laryngeal dysfunction. Clinical utility was defined as the potential to contribute salient diagnostic information to determine oropharyngeal and laryngeal dysfunction. The secondary aim was to provide recommendations to inform the development of SLT-led US protocols and make suggestions for further research for its use in swallowing and laryngeal assessment.

Methods

This review was conducted by a group of eight acute hospital-based SLT clinical experts in response to the request for our national professional body (Royal College of Speech and Language Therapists—RCSLT) to provide a statement on the current utility of US as a swallowing and laryngeal clinical assessment tool. Group membership comprised clinical academics who represented a range of patient populations and geographical regions.

A rapid review was conducted to locate primary research studies using US to assess swallowing and the laryngeal function. The review was based on the methodology and guidance for the conduct of rapid reviews developed by the National Collaborating Centre for Methods and Tools (Dobbins 2017).

Search strategy

Subject and methodological expertise, plus a scoping search of current literature, informed the search strategy. The published literature was identified via an electronic database search of: AMED <1985 to May 2020>, Ovid Emcare <1995 to 2020 week 21>, CINAHL and Medline Complete. Date limits were set for the period January 2010–May 2020 with final searches for all databases completed on 28 May 2020. The following concepts were searched using free text in the title and abstract: ultraso*, sonograph*, ultrasonograph*, dysphag*, swallow*, deglut*, 'pulmonary aspiration', 'respiratory aspiration', 'silent aspiration', 'aspiration pneumonia', tongue, pharynx, larynx, laryngeal, 'vocal cord*', 'vocal fold*', 'vocal ligament*', stridor, bolus (oral OR pharyngeal) AND residue*. In addition, the concepts were mapped to thesaurus subject terms across databases: ultrasonography+, 'deglutition disorders+', 'pneumonia, aspiration', 'respiratory aspiration', tongue+, pharynx+, larynx, 'vocal cords'.

Reference lists of included papers, other relevant reviews and background articles were scrutinized for additional citations. Experts with published work in the area were consulted and electronic alerts for key journals were set up to identify work published after 28 May 2020.

Inclusion and exclusion criteria

Review criteria were designed to reflect the broad scope but short timescales of the rapid review. A population, intervention, comparison, outcome (PICO) framework (Schardt *et al.* 2007) was used to identify primary studies of adults (population) who had undergone US assessment (intervention) alongside a reference test (VFSS, FEES, EEL or validated clinical assessment tool, clinician and/or patient-reported outcome measure) (comparison) where measurement of laryngeal function or swallowing (outcome) had been undertaken. For the purposes of the review, EEL was taken to include direct laryngoscopy (DL), flexible laryngoscopy (FL) or videolaryngoscopy (VL). Database filters were applied to include only English language and exclude papers with non-human participants and those using US to diagnose cancer. Studies that used novel or non-routine comparison tests, such as computed tomography, manometry and muscle biopsy, were excluded as were papers that used US to assess head and neck structure, speech, mastication, intubation and extubation. Any papers with potential clinical utility within SLT but outside the scope of this review were collated as supplementary material.

Selection of publications for review

Database citations were downloaded to Rayyan Qatar Computing Research Institute (QCRI) systematic review web application (Ouzzani *et al.* 2016). Citations were divided into four equal pools and each pool allocated to one of four reviewers (JA, CS, CG and JH). Each reviewer screened their pool at title and abstract level and allocated to one of three predetermined options: 'include,' 'exclude' or 'maybe' based on meeting the PICO criteria. Criteria were refined through iterative discussion to allow resolution of all papers classified as 'maybe.' A fifth reviewer (SW) randomly sampled 10% of each pool for accuracy and any disagreements settled by an additional reviewer (RG).

Five reviewers (JA, CS, JH, CH and GC) used a bespoke data extraction form on two full-text papers as part of a pilot process to discuss and agree standards for data extraction. Data extracted included: primary author and year of publication; country of origin and setting; study design; population and sample size; index and reference test detail; protocol and reliability information as well as key outcomes and findings. Where data formed a section of a multi-part study, only data from the included sub-study were extracted. Full texts were divided between the five reviewers and assessed; papers were excluded from further analysis if they did not meet inclusion criteria.

Critical appraisal

Final full-text papers were assessed for quality using the QUADAS-2 tool (Whiting *et al.* 2011), which assesses four key domains including patient selection, index test, reference standard, and flow of patients through the study and timing of the index test(s) and reference standard. Each study was scored (high, low or unclear) across the four domains. Applicability to a population was scored based on the first three domains only. An a priori decision was made to judge applicability of US as the index test unclear for all papers. This was due to the lack of consensus in the literature around standard test conduct and interpretation for swallowing and laryngeal function. The tool was piloted on one paper by all five reviewers and criteria refined through discussion and consensus. Swallowing papers were assessed by JA, CMG and GC and laryngeal function papers by CS, JH, GC and JA.

Analysis and synthesis of the data

Information from included studies was summarized using tools and techniques of narrative synthesis (Popay *et al.* 2006). This included textual description, grouping and clustering. Visual illustration of findings to show sensitivity, specificity and associated confidence intervals was used where indicated. For studies in which values were missing but sufficient raw data were reported, confidence intervals were calculated using an online calculator http://vassarstats.net/. These studies are identifiable in the summary of included studies (table 1).

Quality assessment findings from QUADAS-2 were summarized into a table by one reviewer with expertise in both swallowing and laryngeal function (GC). Three reviewers agreed a predefined quality scoring system (GC, JH and CS) with final agreement by the first author (JA). High, low or unclear scores for risk of bias and applicability concerns were given to each study based on this system.

Results

Database searching resulted in a total of 2326 papers, with 11 additional records identified through other sources. Deletion of duplicates, abstract and full-text screening resulted in 23 primary studies for inclusion in the final review. The Preferred Reporting Items for Systematic Reviews and Meta-Analyses (Moher *et al.* 2009) flow diagram summarizes the search results and reasons for full-text exclusion (figure 1).

Types of studies, setting and context

An overview of study design, setting and context is provided in table 1. Ten out of the 23 studies were associated with swallowing, and 13 with laryngeal function. The first four sections of table 1 summarize the swallowing studies, listed in order of the oral phase (tongue movement, $n = 1$), pharyngeal phase (hyolaryngeal movement, $n = 4$ and posterior pharyngeal wall movement, $n = 2$) and swallowing symptoms (residue $n = 2$, penetration/aspiration $n = 1$). The final two sections summarize the laryngeal studies which include vocal fold ($n = 12$) and vocal fold plus oedema ($n = 1$) studies.

Swallowing studies originated from East Asia (Japan, $n = 4$; Korea, $n = 2$; Taiwan, $n = 1$; and Hong Kong, $n = 1$) and Italy ($n = 2$), while laryngeal studies were more geographically diverse (Hong Kong, $n = 3$; South Korea, $n = 1$; India, $n = 3$; United States, $n = 3$; Italy, $n = 1$; Spain, $n = 1$; and Egypt, $n = 1$).

Table 1. Characteristics of included studies

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Oral phase swallowing studies									
Tamburrini *et al.* (2010) Italy, MND referral centre	PO	$n = 9$	MND Mean (range) disease duration 15 (6–33) months 44% female; age 60 (33–76) years	ProFocus US system (B-K Medical) 5 MHz micro-convex probe (type 8803) and direct video-capturing software of 25 frames/s with option to slow and freeze	VFSS to define: bolus position in mouth inability to retain bolus in oral cavity reduced and disorganized tongue movement fragmented swallowing pooling of ingested material	Yes Index and reference	No	Six parameters: tongue atrophy abnormal bolus position inability to retain bolus in oral cavity reduced and disorganized tongue movement fragmented swallowing pooling of ingested material Diagnostic markers described for each measure	US > VFSS for identification of: abnormal bolus position (6/9 vs. 3/9) reduced tongue movement (5/9 vs. 2/9) disorganized tongue movement (3/9 vs. 2/9) fragmented swallowing (6/9 vs. 0/9) US = VFSS for identification of inability to retain bolus in mouth (4/9) US < VFSS for identification of pooling (2/9 vs. 0/9)
Pharyngeal phase swallowing studies (hyo-laryngeal movement)									
Chen *et al.* (2017) Taiwan, hospital	PO	$n = 10$	Mixed patient cohort 0% female; age = 71.8 (54–81) years	Self-designed US 3.5 MHz curvilinear transducer recorded at 30 frames/s	VFSS to measure maximum hyoid displacement before and during swallowing.	Yes Index and reference	Yes Interrater intraclass correlation coefficient (ICC) between two examiners 0.892 ($p < 0.05$). ICC between US and VFSS 0.815 and 0.915 for each researcher ($p < 0.01$)	Hyoid bone displacement: distance between hyoid bone and mandible at rest and during swallowing.	No sig. dif. between results of US and VFSS ($p = 0.437$) ICC between VFSS and US for two researchers = 0.815 and 0.916 ($p < 0.001$)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Cheng *et al.* (2018) Hong Kong, hospital	PO X-S	$n = 40$	Post-radiotherapy nasopharyngeal carcinoma patients ≥ 3 years post-treatment 23% female; mean age = 53.9 years	B-mode submental US portable system (Mindray) 6–14MHz linear transducer	VFSS to define: Anterior hyoid displacement Superior hyoid displacement	Yes Index and reference	Yes Intra-rater agreement for US and VFSS; and interrater agreement for US and VFSS. All values ICC ≥ 0.75 ($p < 0.001$)	Geniohyoid contraction: % increase of coronal cross-sectional area	% increase in cross-sectional area of geniohyoid correlated with anterior hyoid displacement Pearson correlation coefficient 0.42 ($p = 0.008$) No correlation with superior hyoid displacement ($r = 0.27$, $p = 0.09$)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Lee *et al.* (2016) Korea, hospital	PO	$n = 52$	Patients with dysphagia (n = 23 ischaemic stroke; $n = 22$ haemorrhagic stroke; $n = 7$ other) 35% female; age = 61.2 (16.4) years	LOGIQ E9 US (GE Healthcare) 1–5 MHz curved probe	VFSS to define: Penetration (PAS 1–8) Residue (Grades 0–3) Subgroups: Non-aspirators Penetrators Aspirators Plus: No residue $< 10\%$ residue $> 10\% < 50\%$ residue $\geq 50\%$ residue	Yes Index only	No	Hyoid bone displacement: distance between hyoid bone and mandible at rest and during swallowing % displacement = hyoid displacement (mm)/resting distance (mm) × 100	Displacement distance sig. ($p < 0.001$) shorter in penetrators and aspirators group than non-aspirators group % displacement sig. smaller in penetrators ($p = 0.001$) and aspirators ($p <$ 0.001) than non-aspirators % displacement in aspirators sig. smaller than penetrators ($p =$ 0.002) Cuff-off value of 13.5 mm for hyoid displacement (sensitivity 83.9%, specificity 81.0%) and 30.3% hyoid bone % displacement (sensitivity 64.5%, specificity 95.2%) to define non-aspirators vs. penetrators/ aspirators

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
									Mean hyoid bone displacement and % hyoid displacement both sig. smaller for group with $> 10\%$ post-swallow residues in piriform fossae than no residue ($p = 0.001$) and $< 10\%$ residue in the piriform fossae ($p = 0.004$) Sig. dif. in mean hyoid displacement between: No residue and $< 10\%$ group ($p = 0.036$) Sig. dif. in mean hyoid displacement and % hyoid displacement between: No residue and $> 10\%$ group ($p < 0.001$) $< 10\%$ residue and $> 10\%$ residue group ($p = 0.005$)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Picelli *et al.* (2020) Italy, hospital	PO	$n = 19$	Acute stroke ($n =$ 14 ischaemic; $n = 5$ haemorrhagic) Mean (SD) disease duration 2.7 (2.1) days 47% female; age = 71.9 (15.5) years	DC-40 US system (Mindray) 6 MHz linear probe	Gugging Swallow Screen (GUSS): Score 0 (worst performance) to 20 (best performance) Functional Oral Intake Scale (FOIS): Scores 1–6 denotes dysphagia Score 7 denotes no dysphagia	No	No	% degree of hyoid-larynx approximation % displacement = resting distance between hyoid and thyroid – shortest distance between hyoid and thyroid during swallowing/initial resting distance × 100	Sig. direct association between FOIS and hyoid-larynx approximation distance ($p = 0.011$ and $r = 0.571$) and degree (%) ($p = 0.005$ and $r = 0.614$) Sig. direct association between GUSS and hyoid-larynx approximation distance ($p = 0.008$ and $r = 0.590$) and degree (%) ($p = 0.004$ and $r = 0.628$) Sig. dif. between dysphagic and not dysphagic WRT hyoid-larynx approximation distance ($p = 0.013$ and $z = -2.494$) and degree ($p = 0.011$ and $z = -2.531$)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Pharyngeal phase swallowing studies (posterior pharyngeal wall movement)									
Kim *et al.* (2012) Korea, hospital	PO	$n = 26$	Stroke ($n = 18$ ischaemic; $n = 8$ haemorrhagic) Mean (SD) disease duration 3.6 (5.2) months 65% female; age = 60 (13.6) years	ACUSON Antares US system, premium edition (Siemens Medical Solutions) 5.71 MHz electronic convex array transducer (Model CH 6-2), B-mode and M-mode	VFSS to define aspira-tion/penetration (group A) vs. not (group B) VFSS parameters compared with US: pharyngeal transit and delay time (PDT) triggering of pharyngeal swallow valleculae and pyriform residue	Yes Index and reference	No	Lateral pharyngeal wall (LPW) displacement and duration of weak side	LPW displacement smaller in group A (0.51 ± 0.37) than group B (0.94 ± 0.43) but not significant ($p = 0.633$) Group A LPW displacement significantly correlated with: Laryngeal elevation ($r = 0.71$, $p = 0.047$) PDT ($r = -0.78$, $p = 0.021$) Valleculae residue ($r = -0.94$, $p = 0.0001$) No correlation with PTT, triggering of pharyngeal swallow or piriform residue ($p > 0.05$) Group B LPW displacement and duration not sig. correlated with residue in the piriform or valleculae, PTT, triggering of pharyngeal swallow, PDT or laryngeal elevation ($p > 0.05$)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Manabe *et al.* (2018) Japan, hospital setting	PO	$n = 56^*$	Patients with mild oropharyngeal dysphagia ($n = 56$) 54% female; age = 58.0 (13.7) years	Aplio XG US system (Toshiba Medical Systems) 12 MHz linear array transducer	VFSS to define: timing of opening and closing of the upper oesophageal sphincter (UES)	Yes Index and reference	Yes	Parameters: maximal movement distance of posterior pharyngeal wall (CE) wall (mm) CE wall opening time (ms) Duration and velocity of CE wall opening and closing	Sig. positive correlation between duration of CE wall opening on US and duration of UES opening on VFSS ($r = 0.86$, $p < 0.001$)
Studies of swallowing symptoms (residue and aspiration)									
Miura *et al.* (2014) Japan, hospital Outpatient clinic	X-S	$n = 17$	Mixed patient cohort Group 1 ($n = 8$): Aspirators. 0% female; age = 71 (9.2) years Group 2 ($n = 9$): Non-aspirators. 11% female; age = 69 (6.2) years	Portable US M-Turbo (Sonosite) 5–15 MHz linear array transducer	VFSS and FEES Binary assessment of aspiration (presence vs. absence)	Yes Index and reference	No	Aspiration: passage of a hyperechoic object through the VF with movement different from the surrounding structure	Of 42 images US correctly detected: Aspiration in 7/11 images identified via VFSS/FEES Absence of aspiration in 26/31 also not identified via VFSS/FEES Aspiration detection on US = 64% sensitivity; 84% specificity (kappa coefficient 0.66)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Miura *et al.* (2016) Japan, hospital Outpatient clinic	X-S	$n = 9$	Mixed patient cohort ($n = 5$ stroke; $n = 1$ Parkinson's disease; $n = 1$ pneumonia; $n = 1$ amyotrophic lateral sclerosis; $n = 1$ healthy) 11% female; mean age = 70 years	Portable US M-Turbo (Sonosite) 5–15 MHz linear array transducer	FEES Binary assessment of residue (presence vs. absence)	Yes Index and reference	No	Post-swallow pharyngeal residue: proportion of high-echogenicity area to the pyriform sinus and vallecula	19 images from nine participants Detection of pharyngeal post-swallow residue on US = 62% sensitivity; 67% specificity

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Miura *et al.* (2020) Japan, hospital	X-S	$n = 35$**	Mixed patient cohort with dysphagia (n = 35). 26% female; age = 80.4 (10.6) years	Handheld US Sonosite iViz (Fujifilm) 5–10 or 6–13 MHz linear array transducers	FEES Classification for level of residue in pyriform fossae (PF) and valleculae (V): none (no boluses or secretions) mild (> 50% PF or V covered severe (< 50% PF or V covered)	Yes Index and reference	No	Post-swallow pharyngeal residue: proportion of high-echogenicity area to the pyriform sinus and vallecula US to detect residue based on echogenicity at cuff-off points < 0, ≤ 0.05, ≤ 0.1 and ≤ 0.5 representing % of high echogenicity area	Cut-off < 0: PF = 92.0% sensitivity (CI = 86.9–95.5); 71.9% specificity (CI = 59.2–82.4) V = 86.7% sensitivity (CI = 75.4–94.1); 63.6% specificity (CI = 40.7–82.8) Cut-off ≤ 0.05* PF = 87.9% sensitivity (CI = 82.1–92.4); 78.1% specificity (CI = 66.0–87.5) V = 85% sensitivity (CI = 73.4–92.9); 81.8% specificity (CI = 59.7–94.8) Cut-off ≤ 0.1 PF = 79.3% sensitivity (CI = 72.5–85.1); 84.4% specificity (CI = 73.1–92.2) V = 75% sensitivity (CI = 62.1–85.3); 86.4% specificity (CI = 65.1–97.1) Cut-off ≤ 0.5 PF = 21.3% sensitivity (CI = 15.4–28.1); 98.4% specificity (CI = 91.6–100) V = 5% sensitivity (CI = 1.0–13.9); 100% specificity (CI = 84.6–100)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Laryngeal studies (vocal fold movement)									
Amis *et al.* (2012) USA, hospital	PO	$n = 16$	Patients with known vocal fold motion abnormalities or perioperative patients having surgery presenting risk to the recurrent laryngeal nerve ?% female; age range = 18–80 years	US; no details of machine provided High-frequency (unstated range) linear probe	DL Binary assessment (normal vs. impaired)	Yes Index only	No	Correlation of US and DL findings for VF motion Data assessed from US image: Alignment of non-phonating VF in relation to the midpoint between them to assess any supero-inferior pre-existing misalignment Latero-medial and/or supero-inferior movements of the VFs during phonation in relation to the midpoint between them. If noted a supero-medial pull on VFs = 'tent-ing'(abnormal)	Congruent findings on 13/16 US to detect VF motion abnormality: 71% sensitivity (CI = 30.2–94.8) 89% specificity (CI = 50.7–99.4) PPF = 83% NPV = 80%

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Caneiro-Pla *et al.* (2014) USA, hospital	PO	$n = 510$	Preoperative patients due to undergo cervical surgery 85% female; age range = 18–86 years	Multiple US systems and probes	Indirect laryngoscopy Binary assessment (normal vs. impaired) Only $n = 70$ had reference test	Yes Index only	No	Visualization of bilateral VF movement	377/510 visualization of bilateral VF movement In $n = 70$: Sensitivity 100% Specificity 98% Accuracy 99% Visualization greater in females vs. males (83% vs. 17%, $p = 0.0005$) Thyroid cartilage calcification affected visualization vs. no thyroid cartilage calcification (42% vs. 81%, $p = 0.0005$)
Dubey *et al.* (2019) India, hospital	PO	$n = 100$	Patients listed for thyroidectomy 67% female; median (IQR) age = 45 (33–54) years	Micromaxx US system (Sonosite) 6–13 MHz linear array transducer	DL and video laryngoscopy Binary assessment (normal vs. immobile)	Yes Index and reference	Yes Perfect agreement for subjective VC assessment ($\kappa = 1.00$, 95% CI = 1.00–1.00); near perfect for VFDV ($\kappa = 0.9994$, 95% CI = 0.9993–0.9995)	Mobility assessment of VFs and VF displacement velocity (VFDV) assessment using US compared with laryngoscopy assessment Distinction between mobile, impaired, immobile VCs and normal VCs	Correlation of US and DL ($r = 0.93$, $p < 0.0001$) Correlation of US and video laryngoscopy ($r = 0.83$, $p < 0.0001$) Preoperative: $r = -0.32$ 95% CI 0.44 to -0.19 $p < 0.0001$ Postoperative: $r = -0.29$, 95% CI -0.40 to -0.15, $p = < 0.0001$

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Fung *et al.* (2020) Hong Kong, hospital	PO	$n = 65$	Patients undergoing elective neck surgery that may pose risk to one or both recurrent laryngeal nerves 68% female; median (range) age = 57 (46–69) years	Portable US system (General Electric) 4–13 MHz linear probe	DL Grading system: Normal Grade 1 (decreased movement) Grade 2 (absence of movement)	Yes Index only	No	US immediately after endotracheal extubation in the recovery room to define: grade I = normal movement grade II = decreased movement grade III = complete absence of movement	$n = 61$ successful US; 94% feasibility 100% correlation between DL and US; grading of movements 100% correlation Sensitivity 100% (CI = 100% (46.3 – 100) Specificity 100% (CI = 92 – 100) PPV 100% NPV 100% Accuracy 100%
Gambardella *et al.* (2020) Italy, hospital	PO	$n = 396$	Patients diagnosed with benign and malignant thyroid disease (preoperative) 66% female; age = 56.4 (18–82) years	MyLab^{TM} X5 (Esatoe) 7−13 MHz linear probe	DL Binary assessment (normal vs. impaired) based on three manoeuvres	Yes Index and reference	No	US to classify normal vs. impaired VF function based on same parameters of DL assessment Notes made on: movement weakness asymmetry paralysis	Accessibility rate of US = 96% US to detect VF alteration: 96.8% sensitivity (CI = 94.4–98.2) 95.6% specificity (CI = 93–97.3) PPV 65.2% (CI = 60.3–79.9) NPV 99.7% (CI = 98.3–100)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Kandil *et al.* (2016) USA, hospital	PO	$n = 250$	Pre- and postoperative parathyroid and thyroid surgery patients 83.2% female; age = 52.7 (14.3) years	US; no details of machine provided 12 MHz linear transducer	DL Binary assessment (normal vs. impaired)	Yes Index only	No	US assessment of active VF movement, classified as normal or impaired	US to detect VF function preoperative: 53.8% sensitivity CI (0.26 – 0.79) 50.5% specificity (CI = 0.46 – 0.55) 50.6% accuracy PPV 2.8% NPV 97.6% US to detect VF function postoperative: 55.6% sensitivity (CI = 0.35 – 0.74%) 38.7% specificity (CI = 0.34 – 0.43%) 39.6% accuracy PPV 4.9% NPV 91.1%
Kumar *et al.* (2018) India, setting unclear	PO	$n = 65$	Patients undergoing thyroid surgery pre- and postoperative (benign or malignant) 72% female; median (range) age = 44 (23–60) years	Portable US (Sonosite) High-frequency (8–12 MHz) linear probe	DL Binary assessment (normal vs. impaired)	Yes Index only	No	Normal vs. abnormal movement on US. Normal movement defined as symmetrical abductive and adductive motion of true VC during quiet respiration	US to detect VF paralysis: 100% sensitivity (CI = 0.34, 1.00) 93.44% specificity (CI = 0.84, 0.97)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Miguel *et al.* (2017) Spain, hospital	PO	$n = 93$	Patients undergoing total thyroidectomy (pre- and postoperative) 78% female; > 18 years old (no further age statistics reported)	Portable US (Mylab 25 Gold) 5–10 MHz linear probe	DL Binary assessment (normal vs. VF palsy) Normal movement = symmetrical abduction and adduction of true VFs at rest and in phonation VF palsy = decreased/absent movement	Yes Index only	No	Evaluation of the accuracy of immediate postoperative period US to diagnose VF paralysis True positive = decreased/absent VF movement on US and confirmed palsy on DL True negative = normal VF movement on US and confirmed palsy on DL False positive = indications of abnormal VF movement on US and normal cord mobility on DL False negative = no abnormal VF movement on US and decreased/absence mobility on DL	Accessibility rate of US preoperative = 94% ($p = 0.99$) US to detect VF palsy: 66.67% sensitivity (CI = 7.4–100%) 100% specificity (CI99.4–100%) PPV 100% (CI = 75–100%) NPV 98.9% (96.2–100%) Accessibility rate of US postoperative = 93% ($p = 0.99$) US to detect VF palsy: 93.3% sensitivity (95% CI = 77.3–100%) 96.1% specificity (95% CI = 91.2–100%) 82.3% PPV (95% CI = 61.2–100%) 98.6% NPV (95% CI = 95.4–100%)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Shah *et al.* (2019) India, hospital	PO	$n = 45$	Patients pre- and post-thyroidectomy (benign or malignant) 87% female; age = 42.02 (15.1) years	Portable US system (Sonosite) High-frequency (5–10 MHz) linear probe	DL Binary assessment: bilateral mobility unilateral mobility could not be assessed	Yes Index only	No	US mobility assessment of VFs as per DL	US to detect VF palsy: 75% sensitivity (CI = 21 – 99%) 95.1% specificity (CI = 85.2 – 99.8%) PPF = 60% NPV = 97.5%
Woo *et al.* (2017) South Korea, hospital	PO	$n = 301$	Patient with postoperative thyroidectomy and other neck operations 82% female; median (range) age = 48 (13–81) years	High definition US system (Philips) High (12−5 MHz) and low (9–3 MHz) frequency	DL Grading system: 1 = normal, symmetrical movement 2 = impaired or decreased movement 3 = no movement	Yes Index and reference	No	High- and low-frequency US to score VF mobility using same grading system as DL	High-frequency US to assess VF motion: 88.4% visualization 92.9% sensitivity (97.5% sensitivity CI = 85.3 – 99.8) 86.5% specificity (99.1% sensitivity CI = 96.5 – 99.8) Low-frequency US to assess VF motion: 97.7% visualization 97.6.9% sensitivity (97.6% sensitiv ity CI = 85.9 – 99.8) 96.5% specificity (99.2% specificity CI = 96.8 – 99.8)

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Wong *et al.* (2014) Hong Kong, hospital	CS	$n = 118$	Patients undergoing thyroidectomy Group 1 ($n =$ 51): vocal cord asymmetry. 92% female; median (range) age = 50 (13–83) years Group 2 ($n =$ 118): no vocal cord asymmetry 83.8% female; median (range) age = 51 (19–78) years	Portable US system iLookTM 25 (Sonosite) 5−10 MHz linear transducer	GRBAS scale and voice impairment scale.	Yes Index only	No	Postoperative VF asymmetry detected by US correlates with voice alteration	Group 1 significantly worse GRBAS ‘G’ score (0.24 vs. 0.07, $p = 0.016$), ‘R’ score (0.33 vs. 0.14, $p =$ 0.022) pre- and postoperation, compared with Group 2
Wong *et al.* (2019) Hong Kong, hospital Linked records: Wong *et al.* (2013) ($n = 204$); Wong *et al.* (2015) ($n =$ 581); and Wong *et al.* (2017) ($n = 1000$) The latest and largest cohort was used for analysis	PO	$n = 1196$	Patients undergoing thyroidectomy (or other neck procedure), pre- and postoperative assessment 79% female; median (range) age = 51 (20–84) years	Portable US system iLookTM 25 (Sonosite) 5−10 MHz linear transducer	DL Grading system: 1 = full or normal, movement 2 = impaired or reduced movement in ≥1 VC 3 = no movement in ≥1 VF	Yes Index only	No	US to grade VF movement as per DL Patients dichotomized into normal (grade 1 vs. abnormal grade 2 or 3)	Diagnosis grade of VF movement on US: 85.3% sensitivity (CI = 0.74 – 0.92) 94.7% specificity (CI = 0.92 – 0.95) PPV 47.9% NPV 99.0%

Continued

Table 1. (Continued)

Reference Country and setting	Study design	Sample size	Population descriptionAge (years): mean (range) or mean (SD), unless specified	Index test equipment	Reference test	Protocol described (Yes/No)	Reliability reported (Yes/No)	Outcomes	Findings
Laryngeal studies (vocal fold movement and *oedema)*									
Kamel *et al.* (2020) Egypt, hospital	PO	$n = 90$	Patients scheduled for anterior cervical spine surgery 36.7% female; 31.1% < 50 years	M Turbo US system (Sonosite) 7–10 MHz linear transducer	Rigid laryngoscopy Binary assessment (normal vs. vocal cord paralysis)	Yes Index only	No	Pre- and postoperative VF oedema: Laryngeal air-column difference VF thickness (mm) Pre- and postoperative VF paralysis: VF movement in breathing and phonation	Diagnosis of postoperative VF oedema on anterior US: 88.2% sensitivity (CI = 62 – 98%) 95.1% specificity (CI = 54 – 93%) PPV = 78.9% NPV = 88.2% Diagnosis of postoperative VF oedema on lateral US: 88.2% sensitivity (CI = 62 – 98%) 94.7% specificity (CI = 71 – 99.7%) PPV = 93.7% NPV = 90% Diagnosis of postoperative VF paralysis on anterior US: 86.7% sensitivity (CI = 74 – 92%) 85.7% specificity (CI = 92 – 95%) PPV = 81.25% NPV = 90% Diagnosis of postoperative VF paralysis on lateral US: 100% sensitivity (CI = 74.6 – 100) 100% specificity (CI = 80.7 – 100) PPV = 100% NPV = 100%

Note: PO, prospective observational; X-S, cross-sectional; CS, case series; vs., versus; SD standard deviation; IQR, interquartile range; PPF, positive predictive value; NPV, negative predictive value; CI, confidence interval; MND, motor neurone disease; US, ultrasound; VFSS, videofluoroscopy swallowing study; FEES, fibreoptic endoscopic evaluation of swallowing; DL, direct laryngoscopy; PAS, penetration aspiration scale (reference); VF, vocal fold; sig., significant; dif., difference.
*Subgroup of a larger population that also included $n = 22$ normal controls for a two-part reliability study.
**Subgroup of a larger population that also included $n = 4$ normal controls for a two-part study where part 1 evaluated the scanning method to detect the anatomical points of interest. Underlining indicates confidence intervals that were calculated by the review authors for purpose of synthesis.

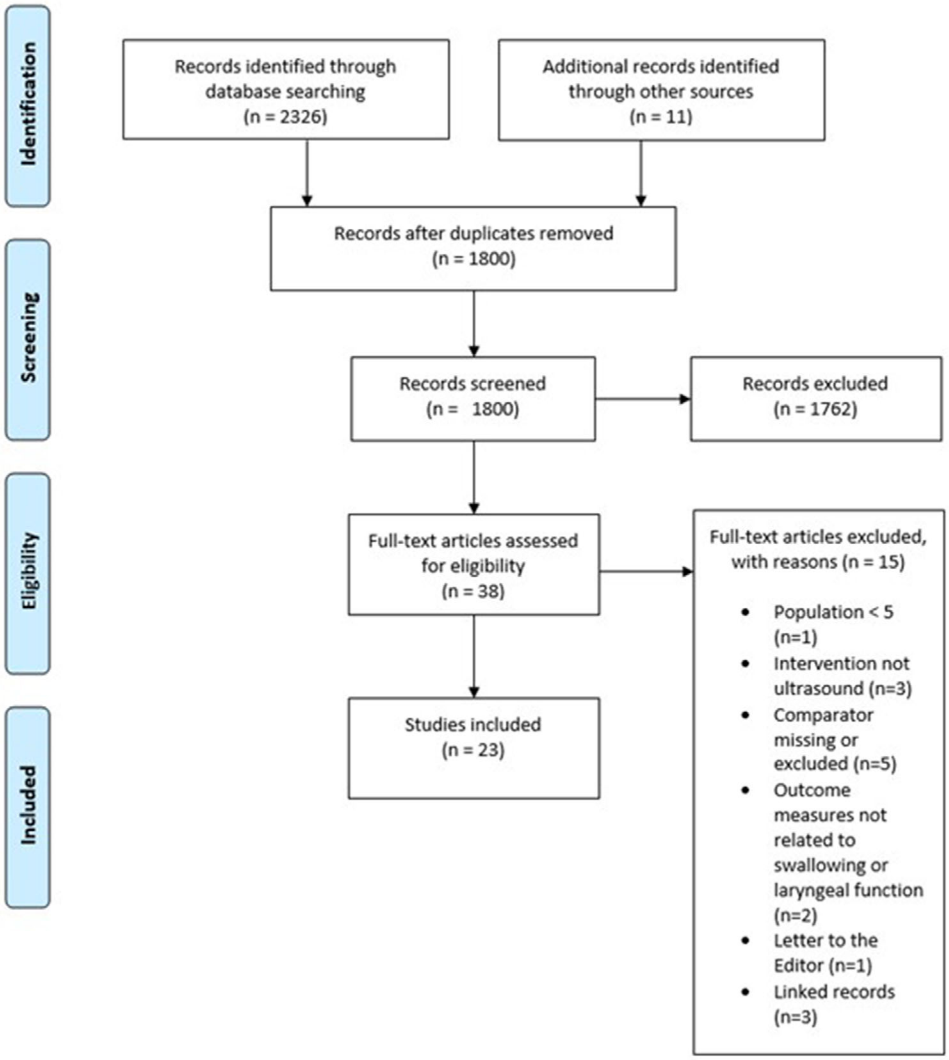

Figure 1. PRISMA flow diagram. [Colour figure can be viewed at wileyonlinelibrary.com]

Studies were prospective observational ($n = 19$), cross-sectional ($n = 3$) and case series ($n = 1$). All except one of the swallowing studies were undertaken in a hospital setting, the remaining being a motor neuron disease referral centre (Tamburrini *et al.* 2010). The laryngeal studies were all conducted in a hospital setting except one where the setting was unclear (Kumar *et al.* 2018).

Study populations

Patient population across the 10 swallowing studies included stroke ($n = 2$) (Kim *et al.* 2012, Picelli *et al.* 2020), motor neuron disease ($n = 1$) (Tamburrini *et al.* 2010), post-radiotherapy ($n = 1$) (Cheng *et al.* 2018) and mixed inpatient cohorts ($n = 6$) (Chen *et al.* 2017, Manabe *et al.* 2018, Lee *et al.* 2016, Miura *et al.* 2014, 2016, 2020). In the 13 laryngeal studies, nine included populations undergoing thyroid surgery ($n = 7$) (Dubey *et al.* 2019, Shah *et al.* 2019, Kumar *et al.* 2018, de Miguel *et al.* 2017, Kandil *et al.* 2016, Gambardella *et al.* 2020, Wong *et al.* 2014), thyroid surgery plus other endocrine-related neck procedures ($n = 1$) (Wong *et al.* 2019) or other neck operations ($n = 1$) (Woo *et al.* 2017). The four remaining studies included participants undergoing neck surgery presenting risk to the recurrent laryngeal nerve ($n = 2$) (Carneiro-Pla *et al.* 2014, Fung and Lang 2020), a mixed neck and vocal fold population ($n = 1$) (Amis *et al.* 2012) and patients undergoing anterior-cervical (AC) spinal surgery ($n = 1$) (Kamel *et al.* 2020).

The 10 swallowing studies had a combined total of 273 participants, with a mean number of 27 and range of 9 (Tamburrini *et al.* 2010, Miura *et al.* 2016) to 56 (Manabe *et al.* 2018). The laryngeal function studies had a total of 3245 participants, with a range of 16 (Amis *et al.* 2012) to 1196 (Wong *et al.* 2019).

Participant gender was reported in all 10 swallowing studies and 11 of the laryngeal studies with 133 female (35.6%) and 240 (64.4%) male, and 2396 female (77%) and 715 males (23%) participants in each respective subgroup.

Participant age was reported in all studies of swallowing and nine (9/13) studies of laryngeal function with a range of 33–91 and 13–86 years, respectively. Mean age of participants across swallowing studies was 65.7 years (SD = 7.82), ranging from 53.9 to 80.4. Mean age of participants across laryngeal studies was 50.5 years (SD = 5.98), ranging from 42 to 58. Papers with a threshold age of <18 years were included as the median age reflected a majority adult cohort.

Ultrasound index test

A range of US equipment was used, including console devices $n = 9$ (Tamburrini *et al.* 2010, Lee *et al.* 2016, Picelli *et al.* 2020, Kim *et al.* 2012, Manabe *et al.* 2018, Dubey *et al.* 2019, Gambardella *et al.* 2020, Woo *et al.* 2017, Kamel *et al.* 2020), portable ($n = 9$) (Cheng *et al.* 2018, Miura *et al.* 2014, 2016, Fung and Lang 2020, Kumar *et al.* 2018, de Miguel *et al.* 2017, Shah *et al.* 2019, Wong *et al.* 2014, 2019), handheld ($n = 1$) (Miura *et al.* 2020), self-made ($n = 1$) (Chen *et al.* 2017) or a combination of multiple systems ($n = 1$) (Carneiro-Pla *et al.* 2014). In all but one swallowing study (Picelli *et al.* 2020) a protocol for conducting the US assessment was reported. Probe and frequency selection varied across the included studies and only four studies provided inter- and intra-rater reliability data (Chen *et al.* 2017, Cheng *et al.* 2018, Manabe *et al.* 2018, Dubey *et al.* 2019).

Reference tests

Six studies (60%) used VFSS to compare US assessment of swallowing biomechanics and/or bolus flow (Tamburrini *et al.* 2010, Goetz *et al.* 2019, Cheng *et al.* 2018, Lee *et al.* 2016, Kim *et al.* 2012, Manabe *et al.* 2018). One study (Picelli *et al.* 2020) used the Gugging Swallow Screen and Functional Oral Intake Scale as a comparator. Two studies (Miura *et al.* 2016, 2020) used FEES to compare with US identification of residue, while the third used a combination of FEES and VFSS (Miura *et al.* 2014) to identify aspiration. A protocol for the reference test is described for all but two of the swallowing papers (Lee *et al.* 2016, Picelli *et al.* 2020). All studies of laryngeal function compared US findings with EEL except one which used a voice impairment scale and GRBAS voice quality perceptual rating (Wong *et al.* 2014). A reference-test protocol was described in only three studies (Dubey *et al.* 2019, Gambardella *et al.* 2020, Woo *et al.* 2017). No studies provided data on rater reliability.

Quality assessment

For a summary of the quality assessment findings, see table 2. Each score (high, low or unclear) is represented symbolically.

Risk of bias

One swallowing paper (Manabe *et al.* 2018) and three laryngeal papers (de Miguel *et al.* 2017, Woo *et al.* 2017, Fung and Lang 2020) had low risk of bias across all four domains. These studies employed consecutive patient selection, appropriate exclusion criteria, blinding and appropriate interval between the index and reference test. Several studies did not recruit consecutive patients and/or patients with potential swallowing or laryngeal difficulties were excluded. Two of the laryngeal studies (Carneiro-Pla *et al.* 2014, Kandil *et al.* 2016) exhibited high risk of bias due to unblinded assessors. Nine out of 10 swallowing studies either did not report or did not employ blinding between reference and index test.

Applicability

All 10 swallowing studies and one laryngeal study (Kamel *et al.* 2020) scored as low for concerns regarding applicability of patient selection. Ten of the laryngeal studies scored high for applicability concerns relating to patient selection. These papers included either a paediatric age range (<18 years) (Wong *et al.* 2014, Woo *et al.* 2017, Dubey *et al.* 2019) or presence of endocrine malignancy in the patient cohort (Woo *et al.* 2017, Dubey *et al.* 2019, Shah *et al.* 2019, Wong *et al.* 2014, 2019, de Miguel *et al.* 2017, Fung and Lang 2020, Carneiro-Pla *et al.* 2014, Gambardella *et al.* 2020). Two papers (Amis *et al.* 2012, Kandil *et al.* 2016) were scored unclear for applicability concerns as they did not provide the diagnosis of participants. All 10 swallowing studies, and all except one of the laryngeal studies scored as low for applicability concerns for choice of reference standard. Wong *et al.* (2014) scored high for applicability concerns as the GRBAS scale was self-rated by patients who provided their own perception of their voice difficulties, despite GRBAS only being validated for clinician assessment.

Summary of the study's findings

Oral phase studies of swallowing

The one identified study of oral phase swallowing function involved a small ($n = 9$) population of patients with motor neurone disease (MND) (Tamburrini *et al.* 2010). Five US parameters of tongue function were compared directly with VFSS measurements. These findings are described in table 1.

Table 2. Quality assessment of studies using QUADAS-2

	RISK OF BIAS				APPLICABILITY CONCERNS		
	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD	FLOW AND TIMING	PATIENT SELECTION	INDEX TEST	REFERENCE STANDARD
Oral phase swallowing studies							
Tamburrini *et al.* 2010	☹	☹	☹	☺	☺	?	☺
Pharyngeal phase (hyolaryngeal movement) swallowing studies							
Chen *et al.* 2016	☹	☹	☹	☺	☺	?	☺
Cheng *et al.* 2018	☹	☹	☹	?	☺	?	☺
Lee *et al.* 2016	☹	☹	☹	☺	☺	?	☺
Picelli *et al.* 2020	?	☹	☹	?	☺	?	☺
Pharyngeal phase (pharyngeal wall movement) swallowing studies							
Kim *et al.* 2012	☹	☹	☹	?	☺	?	☺
Manabe *et al.* 2018	☺	☺	☺	☺	☺	?	☺
Studies of swallowing symptoms							
Miura *et al.* 2014	?	☺	☺	☺	☺	?	☺
Miura *et al.* 2016	☹	?	☺	☺	☺	?	☺
Miura *et al.* 2020	?	☺	☺	☺	☺	?	☺
Laryngeal function (vocal fold movement) studies							
Amis *et al.* 2012	?	☺	☺	?	?	?	☺
Caneiro-Pla *et al.* 2014	☹	☺	☹	☹	☹	?	☺
Dubey *et al.* 2019	☹	☺	☺	☺	☹	?	☺
Fung *et al.* 2020	☺	☺	☺	☺	☹	?	☺
Gambardella *et al.* 2020	?	☺	☺	?	☹	?	☺
Kandil *et al.* 2016	☹	☺	☹	☺	?	?	☺
Kumar *et al.* 2018	?	☺	☺	☺	☹	?	☺
Miguel *et al.* 2017	☺	☺	☺	☺	☹	?	☺
Shah *et al.* 2019	☹	☺	☺	☺	☹	?	☺
Woo *et al.* 2017	☺	☺	☺	☺	☹	?	☺
Wong *et al.* 2014	☹	☺	☹	☺	☹	?	☹
Wong *et al.* 2019	☹	☺	☺	☺	☹	?	☺
Laryngeal function (vocal fold movement and oedema) studies							
Kamel *et al.* 2020	☹	☺	☺	☹	☺	?	☺

Studies of hyo-laryngeal movement

Four studies used US to assess hyo-laryngeal movement. Two measured hyo-laryngeal displacement as defined by the distance between the hyoid bone and mandible at rest and during swallowing (Chen *et al.* 2017, Lee *et al.* 2016), one measured the degree of approximation between the hyoid and larynx (Picelli *et al.* 2020) and the fourth measured geniohyoid contraction by assessing the percentage increase of coronal cross-sectional area (Cheng *et al.* 2018). Chen *et al.* (2017) found no significant differences between measurements of hyo-laryngeal displacement measured by US when compared directly with VFSS. The intra- and interrater reliability and intraclass correlation coefficient (ICC) of the two examiners was found to be excellent as was ICC between US and VFSS (table 1).

Lee *et al.* (2016) used VFSS to estimate aspiration, penetration and residue status after swallowing to establish whether US measurements of hyo-laryngeal displacement can be used to distinguish between clinical groups. Significant differences in hyoid displacement were found between patients with no residue and those with <10% residue and >10% residue ($p = 0.0036$ and <0.001, respectively). A value of 13.5 mm was offered as a cut-off value to distinguish between non-aspirators and aspirators (sensitivity 83.9%, specificity 81.0%). Cheng *et al.* (2017) found that the percentage increase of the geniohyoid cross-sectional area correlated moderately with anterior ($r = 0.42$, $p < .05$) but not superior ($r = 0.27$, $p = 0.9$) hyoid displacement measured by VFSS in 40 post-radiotherapy cancer patients.

Picelli *et al.* (2020) compared degree of hyoid-larynx approximation on US with the Gugging Swallow Screen (GUSS) and Functional Oral Intake Scale (FOIS). Significant differences in hyoid-laryngeal approximation were identified between $n = 19$ dysphagic (FOIS 1–6) versus non-dysphagic (FOIS 7) acute stroke patients. Direct associations were identified between hyoid-laryngeal approximation and FOIS and GUSS scores.

Studies of pharyngeal wall movement

Two studies measured US movement of the posterior pharyngeal wall in stroke ($n = 22$) (Kim *et al.* 2012) and in a mixed ($n = 52$) population (Manabe *et al.* 2018). Kim *et al.* (2012) measured lateral pharyngeal wall displacement of the weak side and compared this with three VFSS parameters described in table 1. In those who aspirated on VFSSS, pharyngeal wall displacement was found to correlate significantly with laryngeal elevation ($r = 0.71$, $p = <.047$), pharyngeal delay time ($r = -0.78$, $p = 0.021$) and valleculae residue ($r = 0.94$, $p < 0.001$). No significant correlations were found between US and VFSS measurements in those that did not aspirate on VFSS.

Manabe *et al.* (2018) measured anterior movement of the posterior pharyngeal (cervical oesophageal—CE) wall and duration and velocity of CE wall opening and closure on US. Significant positive correlations were found between duration of CE wall opening on US and duration of UES opening on VFSS ($r = 0.86$, $p < 0.001$).

Studies of swallowing symptoms

Three studies, all by the same group, assessed the utility of US to detect swallowing symptoms, specifically aspiration (Miura *et al.* 2014) and pharyngeal residue (Miura *et al.* 2016, 2020). Using a binary assessment of residue, Miura *et al.* (2016) found a 62% sensitivity and 67% specificity for use of US as a tool to diagnose residue, which was defined as an 'area of high echogenicity' in the pyriform fossae and/or valleculae. Miura *et al.* (2020) used a more refined method of analysis and provided sensitivity and specificity measures using cut-off points (0%, 5%, 10% and 50%) representing the percentage of a high echogenicity area. A 5% area of high echogenicity provided a superior 87.5% sensitivity (CI = 86.9–95.5) and 78.1% specificity (CI = 40.7–82.8) for diagnosis of pyriform fossae residues and 85% sensitivity (CI = 73.4–92.9) and 81.8% specificity (CI = 59.7–94.8) for diagnosis of valleculae residues. Detection of aspiration by US had a reported 64% sensitivity and 84% specificity when compared with a binary assessment of aspiration on combined VFSS and FEES assessment.

Studies of laryngeal function

Twelve papers compared combined pre- and postoperative sensitivity and specificity of US to measure vocal fold function compared with EEL. Figure 2 provides a visual overview of findings of the included studies.

Sensitivity ranged from 64.3% (Kandil *et al.* 2016) to 100% (Kumar *et al.* 2018, Fung and Lang 2020) and specificity from 48.5% (Kandil *et al.* 2016) to 100% (Kumar *et al.* 2018, Fung and Lang 2020). Visualization of vocal folds was reported in five studies and ranged from 49.1% (Kandil *et al.* 2016) to 100% (Fung and Lang 2020). Figure 2 presents the sensitivity, specificity and confidence intervals for 10 of the 13 studies.

Six studies reported positive predicted value (PPV) (true positives) and negative predictive value (NPV) (true negatives). PPV for US assessment ranged from 47.9% (Wong *et al.* 2019) to 100% (Fung and Lang 2020). NPV for US assessment ranged from 22% (de Miguel *et al.* 2017) to 100% (Fung and Lang 2020). In the six studies, NPV was higher than PPV in three studies (Wong *et al.* 2019, Shah *et al.* 2019, Gambardella *et al.* 2020), while PPV was higher than NPV in one study (de Miguel *et al.* 2017). In one study (Fung and Lang 2020) both PPV and NPV were 100%, indicating perfect positive and negative screening accuracy. Study heterogeneity precluded meta-analysis or any other formal statistical analysis. Three papers were not included in the data synthesis either due to lack of provision of raw data sets (Carneiro-Pla *et al.* 2014, Dubey *et al.* 2019) or differences in reference test (Wong *et al.* 2014).

Caneiro-Pla *et al.* (2014) achieved visualization of the vocal folds in 668/887 patients (77%). Only 70/510 (13.7%) had both EEL and US assessment presenting high risk of bias. Of these, US correctly identified all seven cases with paralysed vocal folds. The sensitivity, specificity and overall accuracy of US in predicting fold paralysis was 100%, 98% and 99%, respectively. Full data sets were not available to calculate confidence intervals. Dubey *et al.* (2019) performed correlation

Pre-operative population	Sensitivity	Lower CI	Upper CI	Specificity	Lower CI	Upper CI
Gambardella et al 2020	96.8	94.4	98.2	95.6	93	97.3
Kandil et al 2016	53.8	26	79	50.5	45	55
Kumar et al 2016	100	34	100	93.4	84	97
Miguel et al 2017	66.7	7.4	100	100	99.4	100
Mixed pre-and post-operative population						
Wong et al 2019	85.3	74	92	94.7	92	95
Post-operative population						
Kandil et al 2016	55.6	35	74	38.7	34	43
Miguel et al 2017	93.3	77.3	100	96.1	91.2	100
Shah et al 2019	75	21	99	95.1	85.2	99.8
Fung et al 2020	100	46.3	100	100	92	100
Post-operative population high vs low frequency (MHz) ultrasound						
Woo et al 2017 (12-5MHz)	97.5	85	99.8	99.1	96.5	99.8
Woo et al 2017 (9-3 MH)	97.6	85.9	99.8	99.2	96.8	99.8
Post-operative population anterior vs lateral placement of ultrasound probe						
Kamel et al 2020 (anterior)	86.7	58	98	85.7	62.6	96.2
Kamel et al 2020 (posterior)	100	74.6	100	100	80.7	100
Mixed surgical and non-surgical population						
Amis et al 2012	71.4	30.2	94.8	88.9	50.7	99.4

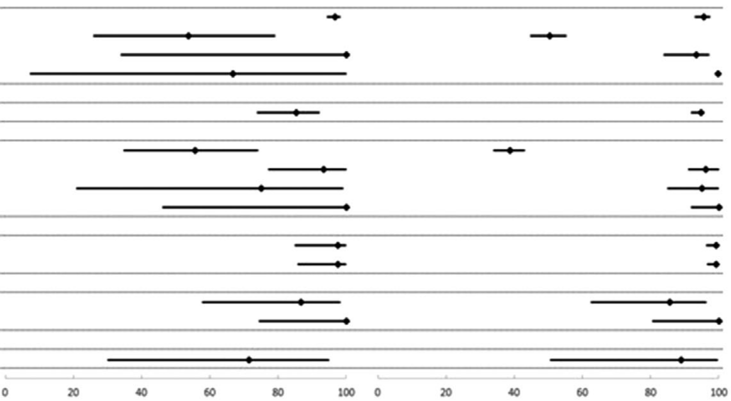

Figure 2. Sensitivity, specificity and confidence intervals of vocal fold studies.

analysis of US versus EEL and found a high correlation ($r = 0.93$, $p < 0.001$) between vocal fold mobility combined with near perfect interrater agreement.

When comparing self-rated GRBAS scale with pre- and post-thyroidectomy US assessment of vocal fold asymmetry, Wong *et al.* (2014) found that participants with vocal fold asymmetry rated themselves significantly higher on the GRBAS 'Grade' score (0.24 versus 0.07, $p = 0.016$) and 'Roughness' score (0.33 versus 0.14, $p = 0.022$) pre- and postoperation, compared with those without asymmetry. Postoperative vocal fold asymmetry detected by US was associated with higher GRBAS scores.

Studies identified a number of factors associated with poorer US visualization of the vocal folds. Age was found to affect US visualization in two studies (Woo *et al.* 2017, Dubey *et al.* 2019), with poorer visualization in older participants. Male gender was associated with poorer visualization (51% compared with 82–96% in females) (Carneiro-Pla *et al.* 2014) and reduced US sensitivity and specificity identified in participants with a higher body mass index (BMI) (Kandil *et al.* 2016). BMI was also highlighted as a non-significant trend by Fung and Lang (2020) but not found to be a significant factor for visualization by Carnierio-Pla *et al.* (2014). Use of low frequency US (3–9 MHz) was found to increase visualization in one paper (Woo *et al.* 2017).

Discussion

This critical review aimed to establish the utility of US as an alternative tool to routine assessments such VFSS, FEES and/or EEL for the clinical assessment of swallowing and/or laryngeal function. It was prompted by the COVID-19 pandemic, but the findings have the potential for application to many patient groups for management of laryngeal function or swallowing. This includes 'hard-to-reach' patient groups where challenges may exist in accessing VFSS, FEES and/or EEL due to geography and/or patient physical and cognitive limitations.

To our knowledge, this is the first comprehensive review of the literature examining the use of US in both swallowing and laryngeal function. We have examined 23 studies that compared US assessment of swallowing or laryngeal function with a standard reference test. All the studies demonstrated a practical ability to visualize structures and the biomechanics of swallowing and laryngeal function using US. However, only two assessed more than one parameter within the same study (Tamburrini *et al.* 2010, Kamel *et al.* 2020). No study combined US assessment of laryngeal function with swallowing, despite the important function of the larynx in airway protection (Pitts 2014).

While there was homogeneity amongst laryngeal studies, in all but one (12/13)study outcome measures were limited to the assessment of vocal fold function in a surgical population. While this restricts the applicability of findings to SLT patients where more complex assessment of laryngeal function is required, the association of vocal fold palsy with glottal incompetence and aspiration (Bhattacharyya *et al.* 2002, Aneas *et al.* 2010, Zhou *et al.* 2018) supports its application to swallowing assessment using US.

Methodological heterogeneity of the swallowing studies prevented in-depth analysis and synthesis of findings. However, this narrative summary has allowed us to expand the findings of the review by Leite *et al.* (2014) progressing our understanding of US as a diagnostic tool for dysphagia and to make future recommendations for application by SLTs.

Swallowing and laryngeal studies

All studies of swallowing biomechanics identified an association or statistical relationships between one or more parameters measured by US compared with VFSS or FEES (Tamburrini *et al.* 2010, Chen *et al.* 2017, Cheng *et al.* 2018, Kim *et al.* 2012, Manabe *et al.* 2018) or between US parameters and clinical surrogates for dysphagia, such as residue, aspiration or restriction in oral intake (Lee *et al.* 2016, Picelli *et al.* 2020). The lack of direct biomechanical relationship between US and VFSS parameters measured by Kim *et al.* (2012) may explain why US measures did not correlate in the group of non-aspirators. The challenges of visualizing areas of high echogenicity instead of anatomical structure and movement may explain the low (<65%) sensitivity of US to detect residue and aspiration (Miura *et al.* 2014, 2016). A more refined method of analysis was used in the later study (Miura *et al.* 2020) leading to a higher (85%) sensitivity.

There is currently no standardized protocol or reference test for US assessment of swallowing. While some studies compare a physiological parameter with the equivalent measure on VFSS and/or FEES imaging, others use surrogate measures for dysphagia such as ratings of residue, aspiration and oral intake scales. The most frequently used parameter for US swallowing assessment was hyoid displacement with agreement of measurement amongst included studies (Chen *et al.* 2017, Lee *et al.* 2016) and existing literature (Chi-Fishman and Sonies 2002, Yabunaka *et al.* 2011, Hsiao *et al.* 2012).

Laryngeal assessment focused on vocal fold mobility rather than other aspects of the larynx, for example arytenoid tilt or vocal fold structure. This simplicity, plus a more standardized approach to assessment amongst pre-existing laryngeal assessment tools may part explain the reasons for the increased homogeneity.

The sensitivity of US to diagnose vocal fold impairment ranged between 63.4% and 100% with a tendency for wide confidence intervals. These figures suggest that the clinical application of US may be best suited as a first-line non-invasive tool to rule out rather than rule in issues with vocal fold mobility. This would correspond with the use of US in other clinical areas (Stengel *et al.* 2018, You-Ten *et al.* 2018).

Kandil *et al.* (2016) had much lower sensitivity and specificity figures than other included studies. This study used a static (12 MHz) frequency probe rather than a spectrum of frequencies (e.g., 6–13 MHz). Lower frequencies are understood to penetrate the larynx more easily allowing for better visualization of structures (Ng and Swanevelder 2011). This was also identified by studies included in the review (Woo *et al.* 2017). The authors propose that participant BMI affected visualization. This is consistent with studies that have associated high BMI with lower quality US images (Brahee *et al.* 2013). Altered body composition may also account for differences in visualization across age (Woo *et al.* 2017, Dubey *et al.* 2019) and gender (Carneiro-Pla *et al.* 2014). Clinicians should be mindful of these challenges when interpreting US findings in these cohorts of patients.

Reference tests

Studies in this review applied a wide range of reference tests. While protocols for these tests were routinely defined, there was poor standardization across studies and infrequent reference to inter- and intra-rater reliability. Absence of reliability reporting is problematic as differences in identification between reference test and US could be considered a simple error. Future studies should use the available standardized and validated scales (Martin-Harris *et al.* 2008, Rosenbek *et al.* 1996, Neubauer *et al.* 2015).

Clinical utility of ultrasound for swallowing and laryngeal assessment

This review has shown that US as a tool for comprehensive swallowing assessment is not currently indicated for use within SLT. However, it does have an emerging role as an assessment of specific structures related to swallowing, including vocal fold mobility. This offers clinical potential as an adjunctive tool. Its role as a complement rather than as a substitute to standard assessments is acknowledged in the wider literature (Fatima *et al.* 2015, Chung and Kim 2015). The unique capability of US to evaluate muscle structure and understand underlying pathology also supports its utility as a supplementary tool (Van Den Engel-Hoek *et al.* 2017).

The absence of protocols in the current literature has impacted on the quality and transferability of the evidence from this review. An important consideration for SLTs is the need for structured training and validated tools in the application and analysis of US findings. Future research is required to promote a standardized approach, including reliability of interpretation and the wider adoption of any tool.

Below we outline some considerations for developing SLT-led protocols and future research studies:

- Visualization and interpretation: Both are operator dependent and require competency development (Pinto *et al.* 2013, Todsen *et al.* 2015, 2018). A major limitation of the included studies was the use of a single-operator design and absent reliability assessment. This limits the reproducibility of the results and may be another explanation for outlying sensitivity and specificity data within the laryngeal studies.
- Equipment: Selection of US machine, probe and frequency varied greatly in this review. Other

clinical areas have acknowledged potential for variation (Aldrich 2007) and have highlighted the need for consensus guidelines, for example, within thyroid assessment (Rago *et al.* 2018). Our review demonstrates that US will need clear standard operating procedures for SLTs to use it as a clinical tool.

- Normative data: The lack of normative data for either vocal fold movement or measurements of structures as a surrogate for swallowing is problematic. Some studies have provided normative values (Miller and Watkin 1997). To identify clinical concern during an US assessment, normative values are necessary.

Strengths and limitations of the review

Multiple individuals participated in the data extraction and quality assessment of this review, increasing potential for interrater differences. To minimize differences methodological safeguards were put in place. These were achieved by developing strict inclusion/exclusion criteria, reassessing a random sample of abstracts, piloting of the data extraction form and group discussion of full text papers, as well as clear written guidelines for the QUADAS-2 assessment.

Due to the pace and context, the team were unable to register the review or publish a formal protocol. Furthermore, as there was no funding attached to the project the remit of the review did not extend to any formal quantitative or meta-analysis. The speed of the review also restricted engagement with patient and public stakeholders who would have been ideally placed to co-develop methodology and provide a unique perspective on the findings. Future research in this area should prioritize the patient perspective of assessment using this tool.

A rapid review methodology was chosen to disseminate findings as quickly as practicable to clinicians globally within the context of the COVID-19 pandemic. Only studies published in English since 2010 were included. Restricting language and date limits may mean that some important studies could have been missed. Extending dates may however have increased variability of findings particularly as US technology has evolved. We had no resources to include studies in other languages.

Future directions

This rapid review has ignited enthusiasm to progress the application of US in the SLT profession via development of clinical protocols for swallowing and laryngeal assessment. This would optimally be combined with training programmes for SLTs to conduct US assessment. These programmes should include identification of key clinical landmarks, static and dynamic assessment techniques, recommendations for equipment selection and cover the technical aspects of operating sonographic equipment. Establishment of inter-rater reliability for key assessment parameters is also an important future goal.

Summary and conclusions

There is emerging evidence to support the utility of US as an adjunct clinical tool for the assessment of swallowing and laryngeal function. Further studies are warranted in a wider range of clinical populations and settings, with increased attention to the enhancement of sensitivity and specificity measures yielded using US. Based on this review, US is currently not recommended as a tool to use in isolation but its potential as a supplementary tool for swallowing and laryngeal assessment is acknowledged. This rapid review has led to an international collaboration that will promote future, targeted research to develop US into a robust and functional clinical tool for use in the management of patients with swallowing and laryngeal difficulties.

Acknowledgements

The authors express their thanks to those in the Ultrasound International Working Group who provided feedback on draft versions of this manuscript. This includes Dr Lenie van den Engel-Hoek, Dr Joan Ma, Dr Katharina Winiker and Professor Maggie-Lee Huckabee. The authors received no specific grant from any funding agency, commercial or not-for-profit sectors to complete this work. Jodi Allen receives funding from the National Institute Health Research (NIHR) for a Pre-Doctoral Research Fellowship. Gemma Clunie receives funding from the NIHR Clinical Doctoral Fellowship Programme. Jemma Haines receives funding from the NIHR Biomedical Research Council for doctoral studies. Roganie Govender receives funding from the NIHR Clinical Lectureship Programme. Supplementary data from this study are available online.

References

Aldrich, J. E., 2007, Basic physics of ultrasound imaging. *Critical Care Medicine*, **35**, S131–S137.

Amis, R. J., Gupta, D., Dowdall, J. R., Srirajakalindini, A. and Folbe, A., 2012, Ultrasound assessment of vocal fold paresis: a correlation case series with flexible fiberoptic laryngoscopy and adding the third dimension (3-D) to vocal fold mobility assessment. *Middle East Journal of Anaesthesiology*, **21**, 493–498.

Aneas, G. C. G., Ricz, H. M. A., Mello-Filho, F. V. and Dantas, R. O., 2010, Swallowing Evaluation in Patients With Unilateral Vocal Fold Immobility. *Gastroenterology Research*, **3**, 245–252.

Bhattacharyya, N., Kotz, T. and Shapiro, J., 2002, Dysphagia and aspiration with unilateral vocal cord immobility: incidence, characterization, and response to surgical treatment. *Annals of Otology, Rhinology and Laryngology*, **111**, 672–679.

Bolton, L., Mills, C., Wallace, S. and Brady, M. C., ROYAL COLLEGE OF, S. & LANGUAGE THERAPISTS, C.-A. G., 2020, Aerosol generating procedures, dysphagia assessment and COVID-19: a rapid review. *International Journal of Language & Communication Disorders*.

Bourke, S. C., 2014, Respiratory involvement in neuromuscular disease. *Clin Med (Lond)*, **14**, 72–75.

Brahee, D. D., Ogedegbe, C., Hassler, C., Nyirenda, T., Hazelwood, V., Morchel, H., Patel, R. S. and Feldman, J., 2013, Body mass index and abdominal ultrasound image quality: a pilot survey of sonographers. *Journal of Diagnostic Medical Sonography*, **29**, 66–72.

Carneiro-Pla, D., Miller, B. S., Wilhelm, S. M., Milas, M., Gauger, P. G., Cohen, M. S., Hughes, D. T. and Solorzano, C. C., 2014, Feasibility of surgeon-performed transcutaneous vocal cord ultrasonography in identifying vocal cord mobility: a multi-institutional experience. *Surgery*, **156**, 1597–1602.

Chen, Y. C., Hsiao, M. Y., Wang, Y. C., Fu, C. P. and Wang, T. G., 2017, Reliability of Ultrasonography in Evaluating Hyoid Bone Movement. *Journal of Medical Ultrasound*, **25**, 90–95.

Cheng, D. T. H., Lee, K. Y. S., Ahuja, A. T. and Tong, M. C. F., 2018, Sonographic assessment of swallowing in irradiated nasopharyngeal carcinoma patients. *Laryngoscope*, **128**, 2552–2559.

Chi-Fishman, G., 2005, Quantitative lingual, pharyngeal and laryngeal ultrasonography in swallowing research: a technical review. *Clinical Linguistics & Phonetics*, **19**, 589–604.

Chi-Fishman, G. and Sonies, B. C., 2002, Kinematic strategies for hyoid movement in rapid sequential swallowing. *Journal of Speech, Language, and Hearing Research: JSLHR*, **45**, 457–468.

Chung, Y. E. and Kim, K. W., 2015, Contrast-enhanced ultrasonography: advance and current status in abdominal imaging. *Ultrasonography*, **34**, 3–18.

Da Costa, B. O. I., Rodrigues, D. D. S. B., Santos, A. S. and Pernambuco, L., 2019, Transcutaneous laryngeal ultrasonography for the assessment of laryngeal function after thyroidectomy: a review. *Ear, Nose & Throat Journal*, 0145561319870487.

De Miguel, M., Peláez, E. M., Caubet, E., González, Ó., Velasco, M. and Rigual, L., 2017, Accuracy of transcutaneous laryngeal ultrasound for detecting vocal cord paralysis in the immediate postoperative period after total thyroidectomy. *Minerva Anestesiologica*, **83**, 1239–1247.

Dobbins, M., 2017, Rapid Review Guidebook: steps for conducting a rapid review. *National Collaborating Centre for Methods and Tools (NCCMT)*.

Dubey, M., Mittal, A. K., Jaipuria, J., Arora, M., Dewan, A. K. and Pahade, A., 2019, Functional analysis of vocal folds by transcutaneous laryngeal ultrasonography in patients undergoing thyroidectomy. *Acta Anaesthesiologica Scandinavica*, **63**, 178–186.

Fatima, S. T., Zahur, Z., Jeilani, A., Hussain, S. J., Abbasi, N. Z., Khan, A. A., Khan, K., Sheikh, A. S., Ali, F. and Memon, K. H., 2015, Ultrasound—A useful complementary tool to mammography in assessment of symptomatic breast diseases. *Journal of Ayub Medical College, Abbottabad*, **27**, 381–383.

Fung, M. M. H. and Lang, B. H., 2020, A prospective study evaluating the feasibility and accuracy of very early postoperative translaryngeal ultrasonography in the assessment of vocal cord function after neck surgery. *Surgery*.

Gambardella, C., Offi, C., Romano, R. M., De Palma, M., Ruggiero, R., Candela, G., Puziello, A., Docimo, L., Grasso, M. and Docimo, G., 2020, Transcutaneous laryngeal ultrasonography: a reliable, non-invasive and inexpensive preoperative method in the evaluation of vocal cords motility—A prospective multicentric analysis on a large series and a literature review. *Updates in Surgery*.

Goetz, C., Burian, N. M., Weitz, J., Wolff, K. D. and Bissinger, O., 2019, Temporary tracheotomy in microvascular reconstruction in maxillofacial surgery: benefit or threat? *Journal of Cranio-Maxillo-Facial Surgery*, **47**, 642–646.

Hsiao, M. Y., Chang, Y. C., Chen, W. S., Chang, H. Y., Wang, T. G., Hsiao, M.-Y., Chang, Y.-C., Chen, W.-S., Chang, H.-Y. and Wang, T.-G., 2012, Application of ultrasonography in assessing oropharyngeal dysphagia in stroke patients. *Ultrasound in Medicine & Biology*, **38**, 1522–1528.

Jones, S. M., Awad, R. E. K., Shaw, J., Slade, S., Stewart, C. and Young, K., 2020, *Speech and language therapy endoscopic evaluation of the larynx for clinical voice disorders: position paper*. London: Royal College of Speech and Language Therapists (RCSLT).

Kamel, A. A. F., Amin, O. A. I., Hassan, M. A. M. M., Elmesallamy, W. A. E. A. and Hassan, E. M., 2020, Ultrasound prediction for vocal cord dysfunction in patients scheduled for anterior cervical spine surgeries: a prospective cohort study. *Journal of Clinical Monitoring and Computing*.

Kandil, E., Deniwar, A., Noureldine, S. I., Hammad, A. Y., Mohamed, H., Al-Qurayshi, Z. and Tufano, R. P., 2016, Assessment of vocal fold function using transcutaneous laryngeal ultrasonography and flexible laryngoscopy. *JAMA Otolaryngology—Head and Neck Surgery*, **142**, 74–78.

Kim, J. H., Kim, M. S., Kim, J.-H. and Kim, M.-S., 2012, Lateral pharyngeal wall motion analysis using ultrasonography in stroke patients with dysphagia. *Ultrasound in Medicine & Biology*, **38**, 2058–2064.

Kumar, A., Sinha, C., Kumar, A., Singh, A., Vardhan, H., Bhavana, K. and Bhar, D., 2018, Assessment of functionality of vocal cords using ultrasound before and after thyroid surgery: an observational study. *Indian Journal of Anaesthesia*, **62**, 599–602.

Lee, Y. S., Lee, K. E., Kang, Y., Yi, T. I. and Kim, J. S., 2016, Usefulness of submental ultrasonographic evaluation for dysphagia patients. *Annals of rehabilitation medicine*, **40**, 197–205.

Leite, K. K. D. A., Mangilli, L. D., Sassi, F. C., Limongi, S. C. O. and Andrade, C. R. F. D., 2014, Ultrassonografia e deglutição: revisão crítica da literatura. *Audiology—Communication Research*, **19**, 412–420.

Manabe, N., Haruma, K., Nakato, R., Kusunoki, H., Kamada, T. and Hata, J., 2018, New ultrasonographic screening method for oropharyngeal dysphagia: tissue Doppler imaging. *American journal of physiology. Gastrointestinal and liver physiology*, **314**, G32–G38.

Martin-Harris, B., Brodsky, M. B., Michel, Y., Castell, D. O., Schleicher, M., Sandidge, J., Maxwell, R. and Blair, J., 2008, MBS measurement tool for swallow impairment–MBSImp: establishing a standard. *Dysphagia*, **23**, 392–405.

Martin-Harris, B. and Jones, B., 2008, The videofluorographic swallowing study. *Physical Medicine and Rehabilitation Clinics of North America*, **19**, 769–785, viii.

McGrath, B. A., Wallace, S. and Goswamy, J., 2020, Laryngeal oedema associated with COVID-19 complicating airway management. *Anaesthesia*, **75**, 972–972.

Miller, J. L. and Watkin, K. L., 1997, Lateral pharyngeal wall motion during swallowing using real time ultrasound. *Dysphagia*, **12**, 125–132.

Miura, Y., Nakagami, G., Yabunaka, K., Tohara, H., Hara, K., Noguchi, H., Mori, T. and Sanada, H., 2016, Detecting pharyngeal post-swallow residue by ultrasound examination: a case series. *Medical ultrasonography*, **18**, 288–293.

Miura, Y., Nakagami, G., Yabunaka, K., Tohara, H., Murayama, R., Noguchi, H., Mori, T. and Sanada, H., 2014, Method for detection of aspiration based on B-mode video ultrasonography. *Radiological Physics and Technology*, **7**, 290–295.

Miura, Y., Yabunaka, K., Karube, M., Tsutaoka, T., Yoshida, M., Matsumoto, M., Nakagami, G., Kamakura, Y., Sugama, J. and Sanada, H., 2020, Establishing a methodology for ultrasound evaluation of pharyngeal residue in the pyriform sinus and epiglottic vallecula. *Respiratory Care*, **65**, 304–313.

Moher, D., Liberati, A., Tetzlaff, J. and Altman, D. G., 2009, Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. *Bmj (Clinical Research Ed.)*, **339**, b2535.

Nakamori, M., Hosomi, N., Takaki, S., Oda, M., Hiraoka, A., Yoshikawa, M., Matsushima, H., Ochi, K., Tsuga, K., Maruyama, H., Izumi, Y. and Matsumoto, M., 2016, Tongue thickness evaluation using ultrasonography can predict swallowing function in amyotrophic lateral sclerosis patients. *Clinical Neurophysiology: official journal of the International Federation of Clinical Neurophysiology*, **127**, 1669–1674.

Neubauer, P. D., Rademaker, A. W. and Leder, S. B., 2015, The Yale Pharyngeal Residue Severity Rating Scale: an anatomically defined and image-based tool. *Dysphagia*, **30**, 521–528.

Ng, A. and Swanevelder, J., 2011, Resolution in ultrasound imaging. *Continuing Education in Anaesthesia Critical Care & Pain*, **11**, 186–192.

Noel, J. E., Orloff, L. A. and Sung, K., 2020, Laryngeal Evaluation during the COVID-19 Pandemic: transcervical Laryngeal Ultrasonography. *Otolaryngology - Head and Neck Surgery*, **163**, 51–53.

Ouzzani, M., Hammady, H., Fedorowicz, Z. and Elmagarmid, A., 2016, Rayyan—A web and mobile app for systematic reviews. *Systematic Reviews*, **5**, 210.

Picelli, A., Modenese, A., Poletto, E., Businaro, V., Varalta, V., Gandolfi, M., Bonetti, B. and Smania, N., 2020, May ultrasonography be considered a useful tool for bedside screening of dysphagia in patients with acute stroke? A cohort study. *Minerva medica*.

Pinto, A., Pinto, F., Faggian, A., Rubini, G., Caranci, F., Macarini, L., Genovese, E. A. and Brunese, L., 2013, Sources of error in emergency ultrasonography. *Critical ultrasound journal*, **5**, **Suppl 1**, S1–S1.

Pitts, T. 2014, Airway protective mechanisms. *Lung*, **192**, 27–31.

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

Popay, J., H., R., Sowden, A., Petticrew, M., Arai, L., Rodgers, M., Britten, N., Roen, K. and Duffy, S., 2006, Guidance on the conduct of narrative synthesis in systematic reviews: a product from the ESRC methods programme. *Economic Social and Research Council.*

Rago, T., Cantisani, V., Ianni, F., Chiovato, L., Garberoglio, R., Durante, C., Frasoldati, A., Spiezia, S., Farina, R., Vallone, G., Pontecorvi, A. and Vitti, P., 2018, Thyroid ultrasonography reporting: consensus of Italian Thyroid Association (AIT), Italian Society of Endocrinology (SIE), Italian Society of Ultrasonography in Medicine and Biology (SIUMB) and Ultrasound Chapter of Italian Society of Medical Radiology (SIRM). *Journal of Endocrinological Investigation*, **41**, 1435–1443.

Rosenbek, J. C., Robbins, J. A., Roecker, E. B., Coyle, J. L. and Wood, J. L. 1996, A penetration-aspiration scale. *Dysphagia*, **11**, 93–98.

Ruan, Z., Ren, R., Dong, W., Ma, J., Xu, Z., Mao, Y. and Jiang, L., 2018, Assessment of vocal cord movement by ultrasound in the ICU. *Intensive Care Medicine*, **44**, 2145–2152.

Schardt, C., Adams, M. B., Owens, T., Keitz, S. and Fontelo, P., 2007, Utilization of the PICO framework to improve searching PubMed for clinical questions. *BMC Medical Informatics and Decision Making*, **7**, 16–16.

Shah, M. K., Ghai, B., Bhatia, N., Verma, R. K. and Panda, N. K., 2019, Comparison of transcutaneous laryngeal ultrasound with video laryngoscope for assessing the vocal cord mobility in patients undergoing thyroid surgery. *Auris, Nasus, Larynx*, **46**, 593–598.

Shawker, T. H., Sonies, B., Hall, T. E. and Baum, B. F., 1984, Ultrasound analysis of tongue, hyoid, and larynx activity during swallowing. *Investigative Radiology*, **19**, 82–86.

Stengel, D., Leisterer, J., Ferrada, P., Ekkernkamp, A., Mutze, S. and Hoenning, A., 2018, Point-of-care ultrasonography for diagnosing thoracoabdominal injuries in patients with blunt trauma. *Cochrane Database of Systematic Reviews (Online)*, **12**, Cd012669.

Suiter, D. M. and Gosa, M. M., 2019, *Assessing and Treating Dysphagia: A Lifespan Perspective*, Thieme Medical Publishers Inc.

Tamburrini, S., Solazzo, A., Sagnelli, A., Del Vecchio, L., Reginelli, A., Monsorrò, M. and Grassi, R., 2010, Amyotrophic lateral sclerosis: sonographic evaluation of dysphagia. *Radiol Med*, **115**, 784–793.

Todsen, T., Melchiors, J., Charabi, B., Henriksen, B., Ringsted, C., Konge, L. and Von Buchwald, C., 2018, Competency-based assessment in surgeon-performed head and neck ultrasonography: a validity study. *Laryngoscope*, **128**, 1346–1352.

Todsen, T., Tolsgaard, M. G., Olsen, B. H., Henriksen, B. M., Hillingsø, J. G., Konge, L., Jensen, M. L. and Ringsted, C., 2015, Reliable and valid assessment of point-of-care ultrasonography. *Annals of Surgery*, **261**, 309–315.

Tran, K., Cimon, K., Severn, M., Pessoa-Silva, C. L. and Conly, J., 2012, Aerosol generating procedures and risk of transmission of acute respiratory infections to healthcare workers: a systematic review. *Plos One*, **7**, e35797.

Van Den Engel-Hoek, L., Lagarde, M. and Van Alfen, N., 2017, Ultrasound of oral and masticatory muscles: why every neuromuscular swallow team should have an ultrasound machine. *Clinical Anatomy*, **30**, 183–193.

Wallace, S., Mclaughlin, C., Clayton, J., Coffey, M., Ellis, J., Haag, R., Howard, A., Marks, H. and Zorko, R., 2020, Fibreoptic Endoscopic evaluation of Swallowing (FEES): the role of speech and language therapy. *A Position Paper. Royal College of Speech and Language Therapists (RCSLT).*

Whiting, P. F., Rutjes, A. W., Westwood, M. E., Mallett, S., Deeks, J. J., Reitsma, J. B., Leeflang, M. M., Sterne, J. A. and Bossuyt, P. M., 2011, QUADAS-2: a revised tool for the quality assessment of diagnostic accuracy studies. *Annals of Internal Medicine*, **155**, 529–536.

Wong, K.-P., Au, K.-P., Lam, S. and Lang, B. H.-H., 2017, Lessons Learned After 1000 Cases of Transcutaneous Laryngeal Ultrasound (TLUSG) with laryngoscopic validation: is there a role of tlusg in patients indicated for laryngoscopic examination before thyroidectomy? *Thyroid: official journal of the American Thyroid Association*, **27**, 88–94.

Wong, K.-P., Au, K. P., Lam, S., Chang, Y. K. and Lang, B. H. H., 2019, Vocal Cord Palsies Missed by Transcutaneous Laryngeal Ultrasound (TLUSG): do they experience worse outcomes? *World journal of surgery*, **43**, 824–830.

Wong, K.-P., Lang, B. H.-H., Chang, Y.-K., Wong, K. C. and Chow, F. C.-L., 2015, Assessing the Validity of Transcutaneous Laryngeal Ultrasonography (TLUSG) after thyroidectomy: what factors matter? *Annals of surgical oncology*, **22**, 1774–1780.

Wong, K.-P., Lang, B. H.-H., Ng, S.-H., Cheung, C.-Y., Chan, C. T.-Y. and Chan, M.-Y., 2014, Is vocal cord asymmetry seen on transcutaneous laryngeal ultrasonography a significant predictor of voice quality changes after thyroidectomy? *World journal of surgery*, **38**, 607–613.

Wong, K.-P., Lang, B. H.-H., Ng, S.-H., Cheung, C.-Y., Chan, C. T.-Y. and Lo, C.-Y., 2013, A prospective, assessor-blind evaluation of surgeon-performed transcutaneous laryngeal ultrasonography in vocal cord examination before and after thyroidectomy. *Surgery*, **154**, 1158–1164.

Woo, J.-W., Park, I., Choe, J. H., Kim, J.-H. and Kim, J. S., 2017, Comparison of ultrasound frequency in laryngeal ultrasound for vocal cord evaluation. *Surgery*, **161**, 1108–1112.

Yabunaka, K., Sanada, H., Sanada, S., Konishi, H., Hashimoto, T., Yatake, H., Yamamoto, K., Katsuda, T. and Ohue, M., 2011, Sonographic assessment of hyoid bone movement during swallowing: a study of normal adults with advancing age. *Radiological Physics and Technology*, **4**, 73–77.

You-Ten, K. E., Siddiqui, N., Teoh, W. H. and Kristensen, M. S., 2018, Point-of-care ultrasound (POCUS) of the upper airway. *Canadian Journal of Anaesthesia*, **65**, 473–484.

Zhou, D., Jafri, M. and Husain, I., 2018, Identifying the prevalence of dysphagia among patients diagnosed with unilateral vocal fold immobility. *Otolaryngology–Head and Neck Surgery*, **160**, 955–964.

Supporting Information

Additional supporting information may be found online in the Supporting Information section at the end of the article.

Supplementary Material