



Using Voice Change as an Indicator of Dysphagia: A Systematic Review

Karoline Weber dos Santos¹ · Esther da Cunha Rodrigues² · Rafaela Soares Rech² ·
Eliana Márcia da Ros Wendland² · Matheus Neves³ · Fernando Neves Hugo³ · Juliana Balbinot Hilgert³

Received: 26 October 2020 / Accepted: 13 May 2021 / Published online: 21 May 2021
© The Author(s), under exclusive licence to Springer Science+Business Media, LLC, part of Springer Nature 2021

Abstract

Voice change post-swallow, observed during a clinical swallow examination, is often used as a clinical indicator of dysphagia risk. However, there has been limited research that evaluated the level of agreement between voice change and swallow dysfunction reported to date. This systematic review aims to investigate existing evidence relating to the relationship between vocal change post-swallow and swallow deficits identified on a Videofluoroscopic Swallow Study (VFSS). The studies were selected by two independent evaluators for inclusion, without restriction on language or date of publication and the methodological quality and the risk of bias were assessed using QUADAS-2. Following the PRISMA recommendation, 271 articles were analyzed, of which 17 were included in the study. Of these, the methodology described in five studies employed voice analyses using only acoustic methods, seven others conducted only auditory-perceptual analyses, and five other studies used both. Across the studies there was no homogeneity in the voice quality parameters assessed, analytic methods used, and results obtained. Forty seven per cent of the studies presented a high risk of bias in the analysis of vocal quality due to lack of clarity and blinding of VFSS. There was no homogeneity in the choice of consistencies evaluated during swallowing, as well as standardization of the outcome investigated in VFSS without a vocal parameter attributable to accurate detection in each outcome. It is not possible to obtain a consensus regarding the recommendation of the use of vocal evaluation as an accurate method for identifying swallowing alterations due to heterogeneity of the vocal evaluation methods, the outcomes evaluated in the VFSS examination, heterogeneity in food and liquid consistencies, and the methodological quality of the studies.

Keywords Deglutition disorders · Deglutition · Voice quality · Voice

Introduction

Oropharyngeal dysphagia may be considered as any alteration to the swallowing capacity characterized by one or more alterations in food and liquid transit from the oral cavity to the esophagus [1, 2]. Videofluoroscopic Swallow Study (VFSS) is classified as the gold standard in the diagnosis of oropharyngeal dysphagia, allowing identification of the

presence of residue along the oropharyngeal tract and the entry of material in the larynx and lower airway, an important risk factor for aspiration pneumonia and malnutrition [1]. Nevertheless, the examination is still difficult to access in some health services and there can be a high cost to the patient. Thus, the clinical evaluation by a specialized professional is the diagnostic method most frequently used in clinical practice [1, 3, 4].

Given the role the larynx plays in protecting the airway during swallowing, and generating sound for voice, some clinical protocols test for pre–post-swallow voice quality change as an indicator of airway protection concerns for persons with dysphagia [3–5]. It is believed that the presence of material retained in the pharyngolaryngeal tract may act as a barrier for the propagation of vocal resonance, given that this change is not observed in individuals with normal swallowing [5, 6]. The characterization of this voice change is variably described in clinical protocols, but typically considers vocal quality parameters such as degree of change

✉ Karoline Weber dos Santos
karolweber@gmail.com

¹ Cristo Redentor Hospital/Conceição Hospital Group, 20, Domingos Rubbo Street, Porto Alegre, Rio Grande do Sul 91040-000, Brazil

² Federal University of Health Sciences of Porto Alegre (UFCSPA), 245, Sarmento Leite Street, Porto Alegre, Rio Grande do Sul 90050-170, Brazil

³ Federal University of Rio Grande do Sul (UFRGS), 110, Paulo Gama, Street, Porto Alegre, Rio Grande do Sul 90040-060, Brazil

(from pre-swallow condition), vocal intensity, and use of the term “wet voice” as a distinct parameter [3, 5, 6]. The term wet voice has been the main term used within the field of dysphagia to characterize change and is described as a sound with a bubbling aspect during post-swallow voicing [4, 7]. Nevertheless, vocal assessment methods for detecting dysphagia are poorly studied and have insufficient evidence for indication as a reliable method.

In the field of voice, a traditional voice evaluation includes auditory-perceptual, aerodynamic, acoustic, and imaging parameters [8]. The auditory-perceptual evaluation is based on training by professionals in the field and quantified using standardized and calibrated measurement scales such as the GRBAS, and the classification of labels such as the attribution of a wet voice [5, 7–9]. In addition, acoustic parameters are used to evaluate physical aspects of sound during phonation [8, 9]. The obtainment of the vocal production for both analytic methods is carried out from the continuous and sustained emission of a vowel, usually /a/, counting of numbers, and reading of texts to obtaining voice samples. This allows for analysis and comparison of voice production variabilities [8], methods also used for the assessment of production variability after swallowing [5–7]. Due to the variability of voicing in the presence of material in the phonatory tract, it is necessary that different assessment approaches with high reliability be tested in order to improve the accuracy of the methods used [4].

Although the evaluation of vocal change is used as one of the indicators of oropharyngeal dysphagia in clinical protocols [3, 4], studies are controversial regarding the diagnostic accuracy relative to the findings of the VFSS [3, 5–7], especially when time linked audio and video data are not available.

The PRISMA-DTA guideline [10] provides guidance for reporting systematic reviews of diagnostic test accuracy studies. When comparing different diagnostic methods, it is important that the diagnostic alternative, named index test, be able to truly identify sick and healthy individuals in a reliable way, when compared to the gold standard, classified as reference standard, and maintaining the same assessment parameters and outcomes investigated.

The aim of this study was to perform a systematic review of published literature examining the relationship between post-swallow voice change and documented swallowing deficits.

Methods

This systematic review was conducted from the guidelines proposed by the Cochrane Handbook for Systematic Reviews of Diagnostic Test Accuracy and reported according to the PRISMA recommendations [10]. This study’s protocol was

published in PROSPERO under register CRD42019131008. The research question of this systematic review was: “Is vocal quality after swallowing an accurate method for detecting oropharyngeal dysphagia?”.

Criteria for Including Studies in This Review

Types of Studies

We included studies that used evaluation methods of phonation compared with results of VFSS, with designs for diagnostic accuracy analysis or observational studies (cross-sectional and cohort studies).

Participants

Patients who underwent evaluations of the swallowing performance from the VFSS, as long as they did not have a tracheostomy in situ or had surgical resections of the phonatory tract structures.

Index Test

For vocal quality evaluation, the vocal sample collection methods from sustained phonation or connect speech performed after swallowing liquid, pasty, or solid food samples were considered. For data analysis, we used acoustic outcomes and auditory-perceptual changes.

Reference Standard

All patients were referred to VFSS for swallowing performance evaluation. Among the variables for conducting the examination, food and liquid consistencies were evaluated, given that swallowing performance may be affected by it [5].

Target Conditions

Listed as outcomes of the VFSS for the characterization of swallowing alterations were the presence of material residues in the hypopharynx, larynx and lower airway.

Search Methods for Study Identification

We conducted a search for studies published in the PubMed/Medline, Cochrane, EMBASE, and Latin American and Caribbean Health Sciences (LILACS) databases, using the search strategies presented in Online Appendix I. Aiming to minimize selection biases, the search was complemented by other bibliographical resources from the health field such as Google Scholar, OpenGrey, ProQuest, dissertations, theses, and reference lists. There was no restriction on language or publication date.

Data Collection and Analysis

Selection of Studies

The studies were initially analyzed by their titles and abstracts by two independent evaluators (KWS and ECR), including studies that met the eligibility criteria. In the cases in which there was no clarity regarding the inclusion, a third evaluator was consulted for discrepancy adjustments (RSR). Those included in this stage were read in full for the final decision on their inclusion.

Data Extraction and Management

After full text reading and inclusion of the studies, the data were extracted by two independent evaluators (KWS and ECR) and checked by a third evaluator (RSR), filling out a data extraction form elaborated for the study. In case of doubts or incomplete data, the authors were contacted. The data extracted were those referring to author, publication year, country, study design, participant age, sample size per gender, and base disease as the cause of oropharyngeal dysphagia. Concerning the outcomes, we collected data on the consistency and volumes of the foods and liquids evaluated, VFSS outcomes analyzed, vocal sample methods, acoustic and auditory-perceptual vocal outcomes, and main results described by the authors.

Assessment of Methodological Quality

The Quality Assessment of Diagnostic Accuracy Studies (QUADAS-2) was used by two independent evaluators (KWS and ECR) to evaluate the quality of the studies. When there was no consensus between the evaluators, a third reviewer (RSR) was consulted.

Results

Study Selection

The described search strategy resulted in a total of 271 studies, of which 73 were duplicated and were excluded. After the removal, 198 titles and abstracts were read by two independent authors, resulting in 27 studies for full reading, without disagreement between the authors, with 17 studies [6, 7, 11–25] finally being included in this review. The PRISMA flow diagram (Fig. 1) presents the selection process and the exclusion reasons at each stage.

Characteristics of the Included Studies

The studies included were published from 1999 to 2018. The sample sizes varied between 6 [11, 22] and 250 [24] individuals. The predominant base diseases in the studies were neurological problems (Table 1).

Videofluoroscopic Swallow Study

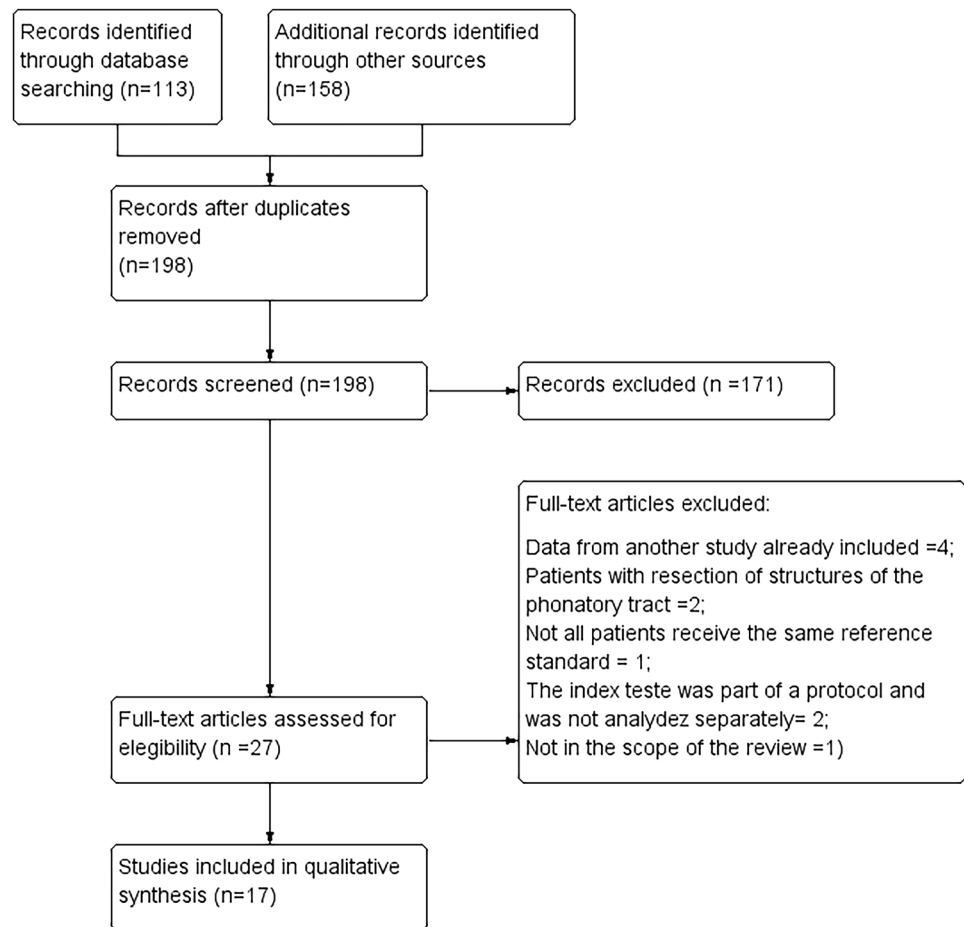
The swallowing function was evaluated from the VFSS, using consistencies of puree [13], liquid [12, 15, 19, 24], or both [6, 7, 11, 14, 16–18, 21, 23, 25]. Only two studies [20, 22] did not report the consistencies used in the evaluations. Considering the number of swallows for analysis, five (29.41%) studies [11, 14, 15, 23, 25] considered one single swallow; three (17.64%) [7, 17, 24], two swallows; three (17.64%) [13, 18, 19], three swallows; and six (35.29%) [6, 12, 16, 20–22] did not report this information. Three (17.64%) studies [7, 15, 19] reported that the video samples were recorded at 30 frames per second and the others did not report this data.

Regarding the outcomes analyzed in the VFSS, only one study [22] did not consider the presence of tracheal penetration or aspiration, only taking into account the gravity of the alteration. Besides examining the presence of tracheal penetration or aspiration, four studies [13, 18, 19, 25] considered the presence of residues in other regions of the pharyngolaryngeal tract for analysis.

Voice Analysis

The vocal production collection method was obtained predominantly from the sustained emission of vowel /a/ in isolation [6, 7, 11, 13, 15, 20, 21, 23–25] or complemented by other emissions [14, 17, 18], with emission from the elevation of pitch [16], non-continuous sound [19], and syllables [22] also being observed. Typically voice samples were collected during sustained phonation /a/ [6, 7, 11, 13–15, 17, 18, 20, 21, 23–25], pitch elevation [16], non-continuous sounds [19], and during production of syllables [22]. In relation to acoustic treatments of the environment in order to reduce noise, one study [6] reported that the vocal samples were obtained in space separated by a soundproof door from the VFSS room; one [7] reported that was not possible to control this aspect; one [19] used an algorithm to remove background noise of the samples; and the others did not report any aspect of controlling.

The vocal samples were obtained before and after swallowing in seven studies [6, 7, 13, 15, 19, 20, 23] to compare the usual vocal pattern to performance after swallowing, while, in the other studies, the vocal samples were only collected after swallowing. In seven (41.17%) studies [11, 12,

Fig. 1 PRISMA flow diagram of the study

[14, 17, 22, 24, 25] the vocal samples were not obtained at the same time as the VFSS. The details are presented in Table 2.

Five studies [6, 15, 20–22] performed vocal analysis from acoustic methods, seven [7, 12, 13, 17, 18, 24, 25] through auditory-perceptual parameters, and five studies [11, 14, 16, 19, 23] used both methods. Six studies (35.29%) [7, 15–18, 22] reported intra or interrater reliability measures for voice analysis. Regarding the quantitative data described in the studies there was no clinical and statistical homogeneity, precluding the data aggregation for meta-analysis. To obtain a comprehensive analysis, we performed a descriptive analysis of the results obtained in each study, presented in Table 2.

Regarding the acoustic analysis parameters investigated, the studies were quite variable concerning the parameters researched, with the most used being the fundamental frequency (f_0) [6, 14–16, 20, 23] and the noise-to-harmonic ratio (NHR) [6, 15, 20]. In six studies [11, 14, 15, 19, 21, 23], no significant vocal changes were identified from the parameters investigated, while three studies [6, 16, 22] identified changes by analyzing different acoustic parameters, yet without accuracy analysis. Only one study identified sensitivity and specificity data over 70% for the relative average perturbation (RAP) and

NHR parameters, identifying an increase in the parameter thresholds after swallowing. However, the data found differ from some included studies that did not present statistically significant results [15] or presented only a reduction in the thresholds of those variables [6].

Among the auditory-perceptual parameters, the items of the GRBAS scale were the main parameters analyzed, having been used partially [11, 13, 14, 17, 25] or completely [19, 23] in the evaluations. There is significant heterogeneity of the investigated variables, with no possibility of data aggregation. Two studies [11, 14] did not identify phonation change after swallowing; two others [13, 23] observed an increase in effort, with sensitivity and specificity around 60% for the presence of residues in the pharyngolaryngeal tract and penetration/aspiration; one study [16] found an increase of pitch after swallowing, yet without accuracy data; and another [25] considering any vocal change, with 72% sensitivity and 67% specificity. Considering the wet voice parameter, nine studies included the variable as an analysis outcome [7, 12, 13, 17, 19, 23–25], while two [11, 14] mentioned the outcome among their results, yet without descriptions in the method sections.

Table 1 Characteristics of included studies

Author and year	Country	Study design	Age	Sample size (F/M)	Cause of oropharyngeal dysphagia
Kang, 2017	Korea	Diagnostic accuracy	Ranged from 30 to 98 years (average 70.8 ± 11.5 years)	165 (74/91)	Stroke: 83 Medical disease: 10 Head-neck cancer: 24 Aging: 35 Degeneration: 13
Groves-Wright, 2010	USA	Diagnostic accuracy	Ranged from 25 to 92 years (average 68.7 ± 13.6 years)	78 (0/78)	Cerebrovascular accident (CVA): 10 CVA and chronic obstructive pulmonary disease (COPD): 4 CVA and gastroesophageal reflux disease (GERD): 5 COPD: 7 COPD and GERD: 7 GERD: 5 Head/neck cancer or injury: 7 Parkinson's disease: 6 Post-cervical spinal surgery: 7 Other neurological disease: 6 Other general medical illness: 14
Vogel, 2018	Germany and Australia	Cross-sectional study	Ranged from 9 to 58 years (average 32.7 ± 13.5 years)	6 (6/0)	Individuals with autosomal recessive spastic ataxia of Charlevoix-Saguenay
Sassi, 2017	Brazil	Diagnostic accuracy	Ranged from 24 to 87 years (average 62.2 ± 16.0 years)	50 (37/13)	20 gastroenterological diseases 15 neurological diseases 8 pulmonary diseases 4 cardiologic diseases 2 vascular diseases 1 rheumatologic disease
Santos, 2017	Brazil	Diagnostic accuracy	Man: average 71.07 Woman: average 76.69	27 (13/14)	22 neurological diseases 5 under clinical investigation
Vogel, 2017	Germany	Cross-sectional study	Ranged from 32 to 71 years (average 50.1 ± 11.2 years)	14 (10/4)	Individuals with mutations in the nuclear-encoded mitochondrial DNA polymerase gamma (POLG)
Chang, 2012	Taiwan	Diagnostic accuracy	Non-P/A Group: 61.8 ± 15.6 P/A Group: 17 (2/15): 66.1 ± 13.4	Non-P/A Group: 27 (9/18) P/A Group: 17 (2/15)	Stroke: 15 Head-neck cancer: 12 Aging: 2 Degeneration: 3 Spinal cord injury: 2 Others: 10
Malandraki, 2011	USA	Diagnostic accuracy	Ranged from 31 to 92 years (average 68 years)	40 (24/16)	Esophageal: 6 Dementia/Parkinson: 7 Stroke: 6 Other neurological conditions: 6 Other medical conditions: 10 Head and neck cancer or trauma: 5
McCullough, 2001	USA	Diagnostic accuracy	Ranged from 40 to 96 years (average 67.8 ± 9.9 years)	60 (5/55)	Stroke
Warms, 2000	Australia	Diagnostic accuracy	Ranged from 20 to 84 years (average 66.6 years)	23 (7/18)	Cerebrovascular accident: 11 Hypoxic brain damage: 1 Motor neuron disorders: 3 Idiopathic: 5 Esophageal motility disorders: 1
Waito, 2010	Canada	Diagnostic accuracy	Not specified	40 (20/20)	Not specified

Table 1 (continued)

Author and year	Country	Study design	Age	Sample size (F/M)	Cause of oropharyngeal dysphagia
Ryu, 2004	Korea	Diagnostic accuracy	Average 64.8 ± 14.3 years	93 (37/56)	Stroke: 43 Medical disease: 20 Diseases and disorders of the ear, nose, throat 12 Parkinson: 6 Meningioma: 4 Dermatomyositis: 3 Others: 5
Ko, 2018	Korea	Cross-sectional study	Average 69.0 ± 11.0 years	30 (14/16)	Parkinson disease: 10 Parkinsonian syndrome: 20
Ryalls, 1999	USA	Diagnostic accuracy	Ranged from 65 to 93 years (average 80.6 years)	6 (0/6)	Not specified for all subjects
Marrara, 2010	Brazil	Diagnostic accuracy	Not specified	31	Neurological disease: 26 Normal individuals: 5
Daniels, 2016	USA	Cross-sectional study	Average 64 ± 9.5 years	250 (7/243)	Stroke 250
Nishiwak, 2005	Japan	Cross-sectional study	Ranged from 30 to 93 years (average 70.4 years)	61 (21/40)	Stroke 61

Most studies [7, 12, 13, 18, 23] did not identify significant changes after swallowing for diagnostic accuracy. Among the studies that conducted contingency analysis [17, 19, 25], low sensitivity (14–50%) was observed, yet with better specificity (78–94%). One of the studies [25] did not allow a specific analysis due to aggregation with other vocal outcomes.

Bias Risk and Evidence Quality of the Included Studies

Figure 2 presents the results of the quality analysis, which is described next.

Patient Selection

Since case–control studies can generate bias in the analysis of data from individuals with alterations, studies with this design were considered to be at high risk of bias in cases where it was not explicitly described that there was blinding analysis as to the diagnosis of swallowing disorders. One (5.88%) study [22] was listed as high risk of bias due to the comparison of individuals with swallowing alterations to healthy individuals without blinded analysis. Regarding the applicability of the data, all studies were considered of low risk, meaning that the profile of the studied patients was in accordance with the objective of the review.

Index Test

As for the risk of bias relative to the vocal evaluation data, five (29.41%) studies [6, 15, 21, 22, 25] presented high bias risks. The high risk occurred primarily due to the absence or lack of clarity as for the blinding of the VFSS data for

vocal evaluation, which may significantly compromise the analysis of the vocal samples. Regarding applicability, all studies were considered low risk since they contemplated the proposal of the review.

Reference Standard

In the risk of bias analysis of the VFSS, only one (5.88%) study [22] was pointed out as high risk due to lack of an appropriate description of the methods for performing the examination. Regarding applicability, all studies were considered low risk because they used the examination as the standard instrument for diagnosis.

Flow and Timing

Vocal emission immediately after swallowing may be changed due to the presence of material residues along the pharyngolaryngeal tract. Thus, the visualization of such residues in the VFSS becomes primordial for the accurate analysis between this finding and the vocal change after swallowing. Based on this, seven (41.17%) studies [11, 12, 14, 17, 22, 24, 25] were classified as high bias risk for not performing the collection of vocal samples immediately after the swallowing observed in the examination.

Quantitative Analysis

Ten studies [7, 12, 13, 16–20, 23–25] presented accuracy data or quantitative details for the composition of contingency tables for accuracy calculations, however due to the qualitative variability and quantitative heterogeneity of the studies, the data were not aggregated in a metanalysis.

Table 2 Diagnostic methods and results obtained by the studies

Author and year	Food consistency and volume assessed	Videofluoroscopy outcomes	Vocal sample method	Sustained vowel /a/	Voice sample collected at the same time as VFSS?	Acoustic outcomes	Auditory-perceptual changes	Reliability measures for voice analysis	Results
Kang, 2017	Yogurt Porridge 5 ml and 10 ml of liquid	Five-point modified penetration aspiration scale	Sustained vowel	/a/	Yes	f0 f0_SD Jitter RAP Shimmer APQ HNR NHR	–	Not described	Reduction of RAP in individuals at risk of aspiration and increase in those at no risk
Groves-Wright, 2010	Pureed Liquid	Presence/absence of penetration/aspiration	Sustained vowel	/a/	Yes	–	Wet voice	Clinicians do not reliably perceive wet voice when material is present in the larynx during phonation	No significant difference for food detection in the pharyngolaryngeal tract
Vogel, 2018	One teaspoon of puree 10 and 30 ml of liquid One bite of bread	Quantitative function across anatomical sites using a four-point scale Penetration–Aspiration Scale	Sustained vowel	/a/	Not clear	Voice quality Vocal control	Voicing Nasality Hoarseness Breathiness Strain	Not described	Voice modification associated with changes in swallowing, but without significance after the adjusted analysis
Sassi, 2017	Videofluoroscopy: 10 ml of liquid Clinical assessment: 5 ml of liquid	Penetration–Aspiration Scale	Not described		24 h after VFSS	–	Wet voice	Not described	Predominance of absence of vocal changes after swallowing (98% of individuals)
Santos, 2017	20 ml of petit suisse yogurt	Presence/absence of: Food stasis in the valleculae Food stasis in the piriform recesses Laryngeal penetration Tracheal aspiration	Sustained vowel	/a/	Yes	–	Grade Roughness Breathiness Asthenia Strain Wet Voice	Not described	Reduction in the Grade and Asthenia, increase in vocal Strain after swallowing, no difference for the wet voice. Asthenia and Strain with sensitivity around 70–80% and specificity lower to 50% to detect penetration or aspiration

Table 2 (continued)

Author and year	Food consistency and volume assessed	Videofluoroscopy outcomes	Vocal sample method	Voice sample collected at the same time as VFSS?	Acoustic outcomes	Auditory-perceptual changes	Reliability measures for voice analysis	Results
Vogel, 2017	One teaspoon of puree 10 ml or 30 ml of water (randomized) Two bites of a dry biscuit	Penetration–Aspiration Scale	Sustained vowel /a/ Saying the days of the week Reading a paragraph Sequential motion rate tasks as quickly and as clearly as possible for 10 s Contemporaneously produced a monologue for 1 min	Not clear	Quantitative measures of timing (mean pause length, percentage of pause, speech rate) vf0 HNR	Hoarseness Breathiness Strain	Not described	Voice modification associated with changes in swallowing, but without significance after the adjusted analysis
Chang, 2012	5-ml bolus of barium sulfate of thin consistency	Presence or absence of penetration/aspiration	Sustained vowel /a/	Yes	f0 RAP Shimmer NHR VTI	–	Not described	No significant difference in vocal parameters
Malandraki, 2011	5 ml of thin liquid from a spoon Thin liquid from a cup (self-administered) 5 ml of a semisolid from a spoon A cookie	3-point residue scoring system Penetration–Aspiration Scale	Pitch elevation task	Yes	Maximum f0 f0 range	Pitch elevation	Acoustic and perceptual voice evaluations had high agreement (95% and 85%, respectively)	Maximum fundamental frequency and pitch increase associated with higher penetration and aspiration scores
McCullough, 2001	5 ml Thin liquid 5 ml Thick liquid 5 ml Puree Solid—1/4 cookie	Presence/absence of aspiration	Sustained vowel /a/ Reading a text	24 h after VFSS	–	Dysphonia Breathiness Harshness Strain Wet voice Resonance	High intra and interjudge reliability according to the authors but without values	Sustained vowel: significance for dysphonia (sensitivity of 100% and specificity of 27%) Text reading: significance for wet voice (50% sensitivity and 78% specificity) and resonance (45% sensitivity and 81% specificity)

Table 2 (continued)

Author and year	Food consistency and volume assessed	Videofluoroscopy outcomes	Vocal sample method	Voice sample collected at the same time as VFSS?	Acoustic outcomes	Auditory-perceptual changes	Reliability measures for voice analysis	Results
Warms, 2000	2.5, 5 and 10 ml of: Thin liquid Thick liquid Pudding	Presence/absence of food: Oral cavity Vallecula Pyriiform sinuses Pharynx Penetration Aspiration	Sustained vowel /a/ Counting from 1 to 5	Yes	–	Wet voice	High levels of agreement (85%)	No association between the presence of a wet voice with penetration or tracheal aspiration
Waito, 2010	5 ml Thin liquid	Penetration–Aspiration Scale Dysphagia: presence of food residues	Phrase ‘ha-ha-ha-ha’	Yes	Jitter Shimmer SNR	Grade Roughness Breathiness Asthenia Strain Wet Voice	Not described	For penetration–aspiration scale: G: sensitivity of 29% and specificity of 75% Wet voice: sensitivity of 14% and specificity of 94% For Dysphagia: G: sensitivity of 25% and specificity of 59% Wet voice: 8% sensitivity and 86% specificity
Ryu, 2004	Not specified	Presence/absence of penetration/aspiration	Sustained vowel /a/	Yes	f0 RAP Shimmer NHR VTI	–	Not described	All parameters increased after swallowing of dysphagic individuals For aspiration or penetration: RAP: sensitivity of 91.1% and specificity of 97.9% NHR: sensitivity of 86.7% and specificity of 77.1%
Ko, 2018	5 and 15 ml Thin liquid 5 and 15 ml Semi-solid	Penetration–Aspiration Scale Dysphagia: presence of food residues	Sustained vowel /a/	Yes	MPT	–	Not described	MPT was not correlated with Penetration–Aspiration Scale

Table 2 (continued)

Author and year	Food consistency and volume assessed	Videofluoroscopy outcomes	Vocal sample method	Voice sample collected at the same time as VFSS?	Acoustic outcomes	Auditory-perceptual changes	Reliability measures for voice analysis	Results
Ryalls, 1999	Not specified	Type/severity of dysphagia	18 syllables	Previously diagnosed by VFSS when allocated to the study	VOT	–	High intra (98%) and interjudge (99%) reliability	For the dysphagic speakers, average VOT values for voiceless stops were shorter, and there were larger negative VOT values for voiced stops. Standard deviations for the VOT productions of the dysphagic subjects were smaller
Marrara, 2010	3 and 7 ml Thin liquid 3 and 7 ml Puree	Presence/absence of penetration/aspiration	Sustained vowel /a/	Yes	Jitter Shimmer VDVP f0	Grade Roughness Breathiness Asthemia Strain Instability Wet Voice	Just one evaluator	Strain and Instability were more sensitive to residue and laryngeal penetration; and for Roughness aspiration. In the acoustic analysis, there was no significant difference in the values between dysphagic and normal
Daniels, 2016	5 ml Thin liquid	Presence/absence of aspiration	Sustained vowel /a/	2 h previously to VFSS	–	Wet voice	Not described	Sensitivity of 27.6% and specificity of 93.2%
Nishiwak, 2005	Videofluoroscopy: 5 ml Thin liquid Teaspoonful Puree Solid Clinical assessment: 5 ml liquid	Aspiration Delayed pharyngeal swallow Pharyngeal residue	Sustained vowel /a/	7 days previously to VFSS	–	Wet voice, Breathiness, Strain or Non-specific hoarseness	Not described	Cough/voice change was significantly associated with aspiration on videofluoroscopic swallow study, with a sensitivity of 72% (95% CI 61–83%) and a specificity of 67% (CI 55–79%)

f0 fundamental frequency, f0_{SD} standard deviation of F0, RAP relative average perturbation, APQ amplitude perturbation quotient, HNR harmonic-to-noise ratio, NHR noise-to-harmonic ratio, v/f0 coefficient of variation of fundamental frequency, VTI voice turbulence index, SNR signal-to-noise ratio, VOT voice onset time, MPT maximum phonation time, VDVP vocal dynamic patterns

Fig. 2 Quality analysis using QUADAS-2

	Risk of Bias				Applicability Concerns		
	Patient Selection	Index Test	Reference Standard	Flow and Timing	Patient Selection	Index Test	Reference Standard
Chang, 2012	+	-	+	+	+	+	+
Daniels, 2016	+	+	+	-	+	+	+
Groves-Wright, 2010	+	+	+	+	+	+	+
Kang, 2017	+	-	+	+	+	+	+
Ko, 2018	+	-	+	+	+	+	+
Malandraki, 2011	+	+	+	+	+	+	+
Marrara, 2010	+	+	+	+	+	+	+
McCullough, 2001	+	+	+	-	+	+	+
Nishiwak, 2005	+	-	+	-	+	+	+
Ryalls, 1999	-	-	-	-	+	+	+
Ryu, 2004	+	+	+	+	+	+	+
Santos, 2017	+	+	+	+	+	+	+
Sassi, 2017	+	+	+	-	+	+	+
Vogel, 2017	+	+	+	-	+	+	+
Vogel, 2018	+	+	+	-	+	+	+
Waito, 2011	+	+	+	+	+	+	+
Warms, 2000	+	+	+	+	+	+	+

High
 Unclear
 Low

Thus, aggregate estimates of accuracy were not reported, while the other parameters of evaluation of diagnostic compatibility, qualitative synthesis of primary studies and descriptive data reported were by the authors following the PRISMA-DTA guideline [10].

Discussion

Most of the studies included in this review investigated the change of vocal production, having as outcomes the

use of the penetration–aspiration scale that graduates the presence and responsiveness of the individual to materials that enters the laryngeal airway and/or is aspirated into the lower airway in eight alteration levels [26]. The presence of laryngeal penetration is characterized by there being material contact with the vocal folds without entering the lower airway, while tracheal aspiration is characterized by material entering the airway, with both responses considering the responsiveness of individuals to such foreign bodies. Although these two aspects present distinct alteration levels and functional impacts, most of the studies performed the analysis of the presence of tracheal penetration or aspiration in aggregate, not allowing precise analyses of what vocal change after swallowing may objectively mean.

The absence of a consensus in terms of aspects to be investigated does not allow an accurate evaluation of the vocal parameters investigated. Moreover, it is known that the viscosity and volume of the food and liquid offered present different neuromotor demands that modify the swallowing performance and, therefore, must be considered for the analysis of each outcome [3, 4, 27]. In view of such aspects, the investigation methods are quite heterogeneous to conduct evaluations capable of identifying one or more swallowing function alterations in an accurate and standardized manner from vocal change after swallowing, presenting high clinical heterogeneity in the attempt of clustering the data, confirmed by statistical heterogeneity.

In addition to the absence of standardization of the objectives of the clinical investigations, it was observed that there was no consensual protocol regarding the voice investigation method. Phonation must be observed before and after swallowing to compare if there is a change to the usual pattern, and it is also essential that this marker is observed at the same time of vocal investigation in order to detail the diagnostic accuracy [5, 26], measures that are little adopted in most studies included in this review, increase the risk of study bias. It is known that the dysfunctions observed in each swallowing may not occur homogeneously at all events [27], and, to be properly associated with a vocal change, both alterations must be investigated simultaneously. Such aspects become essential in the accuracy analysis yet were infrequent across these studies.

The vocal evaluation methods, as well as the parameters used, are also not homogeneous among studies. In the investigation of acoustic data, most did not identify parameters that present significant record variability associated with swallowing, and, among those that did, there is a divergence of results [6, 20]. Among the auditory-perceptual parameters, it is observed that phonation is the most homogeneous indicator that points to the presence of material along the phonatory tract after swallowing [13, 17, 23]; however, there are few studies and with low accuracy for the indication as an isolated diagnostic method, also limited by the

heterogeneity of the reference examination parameters, as described previously. Particularly regarding the wet voice parameter, there is no standardization for the classification, measurement, and analysis of the outcome, with evidence not being observed with association with the parameters that indicate swallowing alteration [7, 12, 13, 18, 23], with a low probability of detecting individuals with alterations (low sensitivity). However, wet voice was highly specific, detecting individuals without the condition [24].

Some study limitations made it impossible to obtain a consensus regarding the recommendation of the use of vocal evaluation as an accurate method for identifying swallowing alterations, including heterogeneity of the vocal evaluation methods, the outcomes evaluated in the VFSS, heterogeneity in food and liquid consistencies, and the methodological quality of the studies; therefore, the vocal evaluation is not recommended as a reliable parameter of the data obtained in VFSS in the identification of swallowing alterations and dysphagia.

The use of parameters of vocal change after swallowing must be analyzed with caution in clinical protocols to avoid measurement biases. Moreover, there is a need to standardize investigation methods for the adequate accuracy analysis of the parameters investigated, especially those of clinical nature, such as the ones evaluated in this systematic review.

Conclusion

It is not possible to obtain a consensus regarding the recommendation of the use of vocal evaluation as an accurate method for identifying swallowing alterations due to the heterogeneity of the vocal evaluation methods, the outcomes evaluated in the VFSS examination, heterogeneity in food and liquid consistencies, and the methodological quality of the studies.

Supplementary Information The online version contains supplementary material available at <https://doi.org/10.1007/s00455-021-10319-y>.

Declarations

Conflict of interest The authors declare that they have no conflict of interest.

References

1. Philpott H, Garg M, Tomic D, Balasubramanian S, Sweis R. Dysphagia: thinking outside the box. *World J Gastroenterol.* 2017;23(38):6942–51. <https://doi.org/10.3748/wjg.v23.i38.6942>.
2. Takizawa C, Gemmell E, Kenworthy J, Speyer R. A systematic review of the prevalence of oropharyngeal dysphagia in stroke, Parkinson's disease, Alzheimer's disease, head injury, and

- pneumonia. *Dysphagia*. 2016;31(3):434–41. <https://doi.org/10.1007/s00455-016-9695-9>.
3. Burgos R, Bretón I, Cereda E, Desport JC, Dziejewski R, Genton L, Gomes F, Jesús P, Leischker A, Muscaritoli M, Poulia KA, Preiser JC, Van der Marck M, Wirth R, Singer P, Bischoff SC. ESPEN guideline clinical nutrition in neurology. *Clin Nutr*. 2018;37(1):354–96. <https://doi.org/10.1016/j.clnu.2017.09.003>.
 4. Murugappan M, Boyce S, Khosla S, Kelchner L, Gutmark E. Acoustic characteristics of phonation in “wet voice” conditions. *J Acoust Soc Am*. 2010;127(4):2578–89. <https://doi.org/10.1121/1.3308478>.
 5. Santos KW, Scheeren B, Maciel AC, Cassol M. Vocal variability post swallowing in individuals with and without oropharyngeal dysphagia. *Int Arch Otorhinolaryngol*. 2015;19(1):61–6. <https://doi.org/10.1055/s-0034-1394129>.
 6. Kang YA, Kim J, Jee SJ, Jo CW, Koo BS. Detection of voice changes due to aspiration via acoustic voice analysis. *Auris Nasus Larynx*. 2018;45(4):801–6. <https://doi.org/10.1016/j.anl.2017.10.007>.
 7. Groves-Wright KJ, Boyce S, Kelchner L. Perception of wet vocal quality in identifying penetration/aspiration during swallowing. *J Speech Lang Hear Res*. 2010;53(3):620–32. [https://doi.org/10.1044/1092-4388\(2009/08-0246\)](https://doi.org/10.1044/1092-4388(2009/08-0246)).
 8. Patel RR, Awan SN, Barkmeier-Kraemer J, Courey M, Deliyski D, Eadie T, Paul D, Švec JG, Hillman R. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. *Am J Speech Lang Pathol*. 2018;27(3):887–905. https://doi.org/10.1044/2018_AJSLP-17-0009.
 9. Stráník A, Čmejla R, Vokřál J. Acoustic parameters for classification of breathiness in continuous speech according to the GRBAS scale. *J Voice*. 2014;28(5):653.e9–653.e17. <https://doi.org/10.1016/j.jvoice.2013.07.016>.
 10. McInnes MDF, Moher D, Thombs BD, McGrath TA, Bossuyt PM, Clifford T, et al. Preferred reporting items for a systematic review and meta-analysis of diagnostic test accuracy studies: the PRISMA-DTA Statement. *JAMA*. 2018;319(4):388–96. <https://doi.org/10.1001/jama.2017.19163>.
 11. Vogel AP, Rommel N, Oettinger A, Stoll LH, Kraus EM, Gagnon C, Horger M, Krumm P, Timmann D, Storey E, Schöls L, Synofzik M. Coordination and timing deficits in speech and swallowing in autosomal recessive spastic ataxia of Charlevoix-Saguenay (ARSACS). *J Neurol*. 2018;265(9):2060–70. <https://doi.org/10.1007/s00415-018-8950-4>.
 12. Sassi FC, Medeiros GC, Zilberstein B, Jayanthi SK, de Andrade CRF. Screening protocol for dysphagia in adults: comparison with videofluoroscopic findings. *Clinics*. 2017;72(12):718–22. [https://doi.org/10.6061/clinics/2017\(12\)01](https://doi.org/10.6061/clinics/2017(12)01).
 13. Santos KWD, Scheeren B, Maciel AC, Cassol M. Postswallow voice modification: compatibility with videofluoroscopy findings. *Codas*. 2017;29(6):e20170004. <https://doi.org/10.1590/2317-1782/20172017004>.
 14. Vogel AP, Rommel N, Oettinger A, Horger M, Krumm P, Kraus EM, Schöls L, Synofzik M. Speech and swallowing abnormalities in adults with POLG associated ataxia (POLG-A). *Mitochondrion*. 2017;37:1–7. <https://doi.org/10.1016/j.mito.2017.06.002>.
 15. Chang HY, Tornig PC, Wang TG, Chang YC. Acoustic voice analysis does not identify presence of penetration/aspiration as confirmed by videofluoroscopic swallowing study. *Arch Phys Med Rehabil*. 2012;93(11):1991–4. <https://doi.org/10.1016/j.apmr.2012.05.028>.
 16. Malandraki GA, Hind JA, Gangnon R, Logemann JA, Robbins J. The utility of pitch elevation in the evaluation of oropharyngeal dysphagia: preliminary findings. *Am J Speech Lang Pathol*. 2011;20(4):262–8. [https://doi.org/10.1044/1058-0360\(2011/10-0097\)](https://doi.org/10.1044/1058-0360(2011/10-0097)).
 17. McCullough GH, Wertz RT, Rosenbek JC. Sensitivity and specificity of clinical/bedside examination signs for detecting aspiration in adults subsequent to stroke. *J Commun Disord*. 2001;34(1–2):55–72.
 18. Warms T, Richards J. “Wet Voice” as a predictor of penetration and aspiration in oropharyngeal dysphagia. *Dysphagia*. 2000;15(2):84–8.
 19. Waito A, Bailey GL, Molfenter SM, Zoratto DC, Steele CM. Voice-quality abnormalities as a sign of dysphagia: validation against acoustic and videofluoroscopic data. *Dysphagia*. 2011;26(2):125–34. <https://doi.org/10.1007/s00455-010-9282-4>.
 20. Ryu JS, Park SR, Choi KH. Prediction of laryngeal aspiration using voice analysis. *Am J Phys Med Rehabil*. 2004;83(10):753–7.
 21. Ko EJ, Chae M, Cho SR. Relationship between swallowing function and maximum phonation time in patients with Parkinsonism. *Ann Rehabil Med*. 2018;42(3):425–32. <https://doi.org/10.5535/arm.2018.42.3.425>.
 22. Ryalls J, Gustafson K, Santini C. Preliminary investigation of voice onset time production in persons with dysphagia. *Dysphagia*. 1999;14(3):169–75.
 23. Marrara JL. Vocal Dynamic Visual Patterns (VDVP) for swallowing dysfunctions analysis in patients with neurological diseases. (2010). M.Sc. Dissertation, Escola de Engenharia de São Carlos, Universidade de São Paulo, São Carlos. <https://doi.org/10.11606/D.18.2010.tde-06052010-160149>.
 24. Daniels SK, Pathak S, Rosenbek JC, Morgan RO, Anderson JA. Rapid aspiration screening for suspected stroke: Part 1: Development and validation. *Arch Phys Med Rehabil*. 2016;97(9):1440–8. <https://doi.org/10.1016/j.apmr.2016.03.025>.
 25. Nishiwaki K, Tsuji T, Liu M, Hase K, Tanaka N, Fujiwara T. Identification of a simple screening tool for dysphagia in patients with stroke using factor analysis of multiple dysphagia variables. *J Rehabil Med*. 2005;37(4):247–51.
 26. Rosenbek JC, Robbins JA, Roecker EB, Coyle JL, Wood JL. A penetration-aspiration scale. *Dysphagia*. 1996;11(2):93–8.
 27. Álvarez G, Dias FJ, Lezcano MF, Arias A, Fuentes R. A novel three-dimensional analysis of tongue movement during water and saliva deglutition: a preliminary study on swallowing patterns. *Dysphagia*. 2019;34(3):397–406. <https://doi.org/10.1007/s00455-018-9953-0>.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Karoline Weber dos Santos M.Sc

Esther da Cunha Rodrigues B.Sc

Rafaela Soares Rech PhD

Eliana Márcia da Ros Wendland PhD

Matheus Neves PhD

Fernando Neves Hugo PhD

Juliana Balbinot Hilgert PhD