A Systematic Review of Speech-Language Pathology Interventions for Presbyphonia Using the Rehabilitation Treatment Specification System

*[1]Brian Saccente-Kennedy, *Fiona Gillies, †Maude Desjardins, ‡Jarrad Van Stan, and §,¶Roganie Govender,
*§¶*London, UK,* †*Québec, Canada, and* ‡*Boston, Massachusetts*

Summary: Background. The prevalence of voice disorders for people aged > 65 years is four times higher than for the population at large. The most common cause of dysphonia in this group is presbyphonia, the preferred first-line treatment for which is voice therapy with a speech-language pathologist. This systematic review seeks to identify how voice therapy affects multidimensional voice outcomes in people with presbyphonia.

Methods. A systematic search of CINAHL, Embase, Emcare, MEDLINE, and Google Scholar was conducted in March 2023. Comparative and noncomparative studies of voice therapy in participants aged > 50 years with presbyphonia were considered for inclusion. No limitations were placed on date or language of publication. Study quality and risk of bias were assessed with the Cochrane *Risk of Bias 2* tool and the *Methodological Index for Non-Randomized Studies*. Subgroup analysis was used to compare studies based on participant sex, intervention duration, study design, and intervention content. Interventions were specified using the Rehabilitation Treatment Specification System (RTSS) employing a consensus methodology among reviewers. The results were synthesized utilizing meta-analysis when outcomes were adequately specified and narrative analysis when they were not.

Results. Twenty-three studies were included with 1050 subjects (mean age: 72.5 ± 8.6 years; 51% female). The most reported intervention was vocal function exercises. Per the RTSS, 14 interventions employed a predominantly *Organ Functions* approach, and the 14 remaining interventions employed a *Skills & Habits* approach. Meta-analysis confirmed posttherapy improvement in patient-related outcome measures of 0.93 standard mean difference (P < 0.00001, 95% confidence interval [CI]: 0.70–1.17); studies with predominantly males and with longer treatment periods were associated with larger improvements, while randomized controlled trials reported more modest improvements. Meta-analysis also identified a mean posttherapy increase in maximum phonation time (MPT) of 5.37 seconds (P < 0.00001, 95% CI: 3.52–7.22). Treatments with an *Organ Functions* focus resulted in greater gains in MPT than those with a *Skills & Habits* focus (7.52 seconds versus 2. 90 seconds). Finally, meta-analysis identified reductions in acoustic perturbation measures (*jitter:* 0.62%, P < 0. 001, 95% CI: 0.26%–0.97%; *shimmer* 1.05%, P < 0.00001, 95% CI: 0.67%–1.44%). Narrative synthesis further identified improvement in auditory-perceptual voice quality in all active treatment groups as well as improved glottal function in most studies that reported this.

Conclusions. Despite the uncertainty around internal validity introduced by the inclusion of a wide range of study designs, there is convincing evidence that voice therapy for presbyphonia results in significant improvement in patient-reported, aerodynamic, acoustic, and expert-rated voice outcomes. Treatments with an *Organ Functions* focus may better address the underlying physiological deficits of presbyphonia, although future comparative studies with multidimensional voice assessment are warranted.

Key Words: Presbyphonia–Aging voice–Vocal fold atrophy–Voice therapy–Speech-language pathology–Speech and language therapy.

INTRODUCTION

The prevalence of voice disorders for adults aged > 65 years is 20%–30%, approximately four times higher than for the population at large and representing over 20% of all dysphonia diagnoses across all ages.[1–3] For those living in care homes, prevalence is even higher at 40%.[4] The proportion of over-65 seconds in the general population will not reach 20% until 2030[5]; however, they already outpace this percentage in referrals to voice centers, placing an increasing strain on health services.[6] Voice disorders in older adults can arise from a variety of etiologies, but dysphonia caused by age-related changes to the larynx and lower respiratory system, termed *presbyphonia,* is the most commonly reported[6,7] and is on the

Accepted for publication December 6, 2023.
From the *Department of Speech and Language Therapy (ENT), Royal National Ear, Nose and Throat and Eastman Dental Hospitals, University College London Hospitals NHS Foundation Trust, London, UK; †École des sciences de la réadaptation, Faculté de médecine, Université Laval, Québec, Canada; ‡Harvard Medical School, Boston, Massachusetts; §University College London, Division of Surgery & Interventional Science, London, UK; and the ¶University College London Hospital, Head and Neck Centre, London, UK
[1] First author..
Address correspondence and reprint requests to Brian Saccente-Kennedy, Department of Speech and Language Therapy (ENT), Royal National Ear, Nose and Throat and Eastman Dental Hospitals, University College London Hospitals NHS Foundation Trust, 47-49 Huntley Street, London, WC1E 6DG, UK. E-mail:
Journal of Voice, Vol xx, No xx, pp. xxx–xxx
0892-1997

https://doi.org/10.1016/j.jvoice.2023.12.010

rise.[8] The mainstay first-line treatment for presbyphonia is voice therapy with a speech-language pathologist.[9] It seeks to reduce atrophy of the vocal folds, coordinate airflow with phonation, optimize resonance, pitch, loudness, and undo maladaptive patterns of laryngeal constriction.[10,11] Although little aside from small case studies were published up until 2008,[11] evidence for the role of voice therapy in this population has been accumulating over the intervening period.

A number of overview articles on the pathophysiology of presbyphonia and its treatment, including voice therapy, are available but have not employed an explicitly systematic methodology.[11,12] In contrast, Marchand and Bonamigo published a systematic exploratory review of the broad application of speech-language therapy in the area of elderly voice,[13] but their search was not confined to presbyphonia nor to voice therapy rehabilitation. Galluzzi and Garavello, and more recently Bhatt et al, have published systematic reviews of surgical and behavioral treatment options for presbyphonia/age-related vocal fold atrophy. In the former,[14] 10 papers involving voice therapy were described in a narrative synthesis; however, no detail was given of eligibility criteria for including studies, and the authors were unable to offer conclusions on the effectiveness of particular approaches. In the latter paper,[15] Bhatt et al used a well-specified systematic methodology to review the quality of the evidence and compare the relative efficacy of surgical and voice therapy interventions. Opting to forego quantitative data synthesis owing to heterogeneity among the studies, the authors called for future studies to use a common, multidimensional voice assessment to enable comparison between interventions. They concluded that both surgical and voice therapy interventions appeared to be superior to control groups but were unable to conclude on the relative superiority of one over the other. Finally, Leão et al published the only systematic review to date focusing solely on voice therapy in an elderly voice population.[16] This narrative synthesis was ostensibly of voice therapy interventions with a respiratory approach. Unfortunately, the authors did not make clear why the included studies were considered to have such a respiratory focus (beyond the fact that they broadly addressed some aspect of breathing during phonation). Their inclusion criteria also allowed for subjects with neurological disease (Parkinson's disease), thereby introducing heterogeneity to the sample.

Although systematic reviews and meta-analyses are considered gold standard evidence *that* a treatment works for a particular population, they offer limited practical information for clinicians in practice.[17] It would be more useful to understand the active treatment ingredients and therapeutic targets, in other words, *how* a therapeutic approach works and on *what* aspects of functioning, so that an intervention can be tailored to a specific patient, with specific needs, in a specific context. Systematic reviews are not typically designed to uncover such elements; however, utilizing a framework to specify the treatment components of an intervention can highlight similarities in therapy ingredients and targets and enable useful comparisons between studies with similar approaches. In this way, similarities or differences between groups of treatments can be evaluated considering outcome measures of interest.

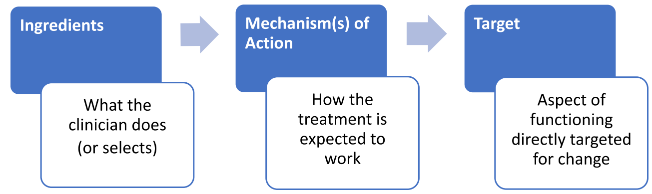


FIGURE 1. RTSS tripartite treatment component (taken from the Manual of Rehabilitation Treatment Specification[19]).

We employed the Rehabilitation Treatment Specification System (RTSS)[18] to describe the interventions utilized by the included studies. The RTSS conceptualizes a tripartite treatment component structure (Figure 1) to predict therapeutic change. One or more clinician-delivered ingredients are theorized to directly affect an element of patient functioning, a therapeutic target, via a proposed mechanism of action.[19]

There are three mutually exclusive groups of treatment components specified by the RTSS: *Organ Functions*, targeting the function/structure of an organ; *Skills & Habits*, targeting the ability to perform a task/habit; and *Representations*, targeting modified knowledge/attitudes or probability of behaviors. Whereas the number of possible *Representations* targets and ingredients is possibly infinite, the RTSS developers envision finite expert curated sets in respect of *Organ Functions* and *Skills & Habits* treatments, from which clinicians might select as if from a menu.[19] Such a consensus list for voice therapy has recently been produced by Van Stan et al, the *RTSS-Voice*;[20] it forms the basis of presbyphonia treatment specification for this review.

Our primary objective was to identify how voice therapy affects multidimensional voice outcomes for people diagnosed with presbyphonia through a broad synthesis of the relevant literature utilizing meta-analysis where possible. A secondary objective was to use subgroup analysis to make tentative comparisons between studies based on participant sex, intervention duration, study design, and intervention content (RTSS treatment groups).

METHODS

We report our methods and findings according to the Page et al's guidelines in the Preferred Reporting Items for Systematic reviews and Meta-Analyses (PRISMA 2020 statement).[21] Please see Supplemental Information for the completed PRISMA 2020 checklist.

Eligibility criteria

Studies were eligible if they met the inclusion and exclusion criteria summarized in Table 1. Presbyphonia is often described as a diagnosis of exclusion,[22] with a symptomatic

TABLE 1.
Study Eligibility Criteria

Inclusion criteria	
Population	Adults aged $\geq$50 years with a diagnosis of presbyphonia/age-related dysphonia/vocal fold atrophy
Intervention	Behavioral voice therapy interventions delivered by/within scope of a speech-language pathologist
Outcomes	Patient-reported voice outcomes, objective/subjective measures of voice quality, and function
Design	Randomized and nonrandomized studies with and without a control group
Exclusion criteria	
Population	Adults aged $<$50 years, concurrent additional laryngeal pathology, or neurological/progressive degenerative disease
Intervention	Neuromuscular electrical stimulation/functional electrical stimulation
Design	Case series/reports where $n \leq 5$
Publication type	Conference posters and conference abstracts

course known to exhibit considerable variation between individuals.[23] The presence of laryngeal aging (presbylarynx), alone, does not guarantee symptomatic presbyphonia,[24] necessitating a wider consideration of other age-related changes to the physiology of voice production.[12] Despite chronological age being a poor predictor for biological aging,[10] a threshold age of 50 years was selected, as this is the age after which noticeable changes to respiratory structure and function occur,[25] affecting the main power source of the voice.

Information sources

The search was conducted in March 2023 of four databases: CINAHL, Embase, Emcare, and MEDLINE. Google Scholar (scholar.google.com) was also searched to identify eligible gray literature or papers missed in the database search. Bibliographic searching was utilized throughout the identification process, and papers were considered for inclusion based on study title, abstract, and full text review.

Search strategy

Our search strategy was based upon subject heading and free text terms representative of the concepts of *voice* + *aging* + *voice therapy* (refer to Appendix A for the full search strategies utilized). No publication date or publication language restrictions were imposed.

Selection process

Eligible citations were imported into Covidence (www.covidence.org), wherein screening decisions, critical appraisal, and data collection were undertaken. Initial title and abstract screening were undertaken by B.S.K. to identify studies to progress to full text eligibility screening. Full texts were retrieved and screened independently by B.S.K. and F.G. against eligibility criteria. Disagreements over eligibility were resolved at a meeting between reviewers. For papers published in a language other than English, full texts were machine translated semiautomatically using translate.google.com, and resultant texts checked for clarity.

Data collection process

Data were extracted independently from the eligible studies by B.S.K. and F.G. utilizing a custom data extraction template on Covidence. Any disagreements or inconsistencies were resolved at a subsequent meeting between reviewers. In the event of unclear or inadequately specified data in the source studies, B.S.K. contacted authors to obtain clarification.

Data items

Data were extracted for voice-related outcomes to speech-language pathology interventions for presbyphonia. These were broadly categorized as patient-related outcome measures (PROMs), objective measures of voice production (aerodynamics, respiratory function, acoustics, objective measures of laryngeal/glottal function), and subjective/expert-rated outcome measures (auditory-perceptual voice quality and subjective measures of laryngeal/glottal function). Data pertaining to intervention characteristics (dosing and duration of treatment and number of clinic visits) and participant characteristics (age and sex) were also extracted.

Study risk of bias assessment

Studies were assessed independently by B.S.K. and F.G. using a custom critical appraisal template on Covidence. For randomized controlled trials (RCTs), the Cochrane *Risk of Bias 2* (RoB) tool was employed.[26] For all other study designs, the *Methodological Index for Non-Randomized Studies* (MINORS) was used.[27] The RoB 2 tool utilizes an algorithm to report three levels of risk of bias (*low*, *high*, and *some concerns*) across five domains for RCTs: randomization process, deviations from intended intervention, missing outcome data, measurement of the outcome, and the selection of the reported result. The algorithm outputs an overall risk of bias. For MINORS, studies are scored against 12 items (or a subset of 8 for noncomparative studies). The scoring system awards 2 points if the item is reported and adequate, 1 point if it is reported and inadequate, and 0 points if it is not reported. This leads to a maximum score of 24 points for a comparative study and 16 points for a noncomparative study.

Any disagreements in risk of bias assessment between reviewers were resolved by discussion.

Effect measures

Outcome data were summarized and tabulated. For studies that adequately reported the spread of data (see next section), bias-corrected Hedge's g (an effect size suited to smaller studies as sample size is considered in the calculation) was determined using the following formula:

$$g = \frac{M_1 - M_2}{SD_{pooled}} \times \left(\frac{N-3}{N-2.25}\right) \times \sqrt{\frac{N-2}{N}}$$

In studies for which an effect size could not be determined, the direction of change and level of significance have been reported instead. When appropriate, outcomes were synthesized in meta-analysis. As these outcomes represented continuous data, mean difference/standard mean difference (SMD) were utilized to report effect measures using a randomized effect model.

Synthesis methods

Participant characteristics and details of the intervention(s) (approaches/therapies, dosing of home practice, duration of intervention) were tabulated from data extraction templates, as were the pre- and post-intervention outcomes. For data that were insufficiently specified for meta-analysis (ie, no measure of the spread of the data in the form of a standard deviation [SD] or interquartile range [IQR]), B.S.K. contacted authors to request these data. For data that reported median/IQR and which were to be included in meta-analysis, the mean/SD were approximated as per the methods proposed by Wan et al.[28] Where the data for two or more study subgroups needed to be combined, means/SD were combined using the Cochrane method.[29] Only in the case of approximating age distribution for all participants across all studies, the range rule[30] was applied to approximate SD in studies that only reported mean and range.

A first-stage result synthesis involved specifying the extracted intervention detail as per the RTSS,[18] guided by the consensus list of voice targets and ingredients provided by RTSS Voice.[20] B.S.K. and F.G. independently reviewed the extracted data and proposed a number of intervention components (each having a singular target and one or more ingredients) for each intervention reported in the included studies. A consensus was arrived at for each treatment component and its corresponding target and ingredient(s) in discussion between the reviewers and J.V.S. The resulting RTSS specification for the interventions included within the studies allowed for the identification of intervention targets and their predominant treatment group (*Organ Functions* or *Skills & Habits,* no studies were characterized by a predominant *Representations* focus) allowing for meaningful subgroup analyses. The predominant treatment group was determined in discussion between B.S.K. and F.G. and reflected the treatment theory (implicit or explicit) of the study authors for the given intervention.

Other subgroup categories that were identified by the reviewers from study and participant characteristics were treatment duration (< 12 weeks versus ≥12 weeks), number of clinician-led sessions (< 6 versus ≥6), study design (RCT versus non-RCT), and sex composition (predominantly males [> 65%] versus mixed sex [< 65% males]). For duration, session number, and sex, thresholds were pragmatically set with the aim of producing groups of sufficient size to enable tentative subgroup comparison. Treatment duration was differentiated from number of sessions to discern the possible influence of home practice dose (treatment duration) as well as clinician input dose (number of sessions).

Our principal research question is to identify how speech-language pathology interventions affect voice outcomes; it is broad and aims to capture a wide snapshot of practices. We have been heuristic in our inclusion of a variety of study designs, despite the challenge to internal validity presented by noncomparative designs. Meta-analysis was conducted only for the most represented, continuous outcomes. This enabled a more robust assessment of dispersion of outcomes and enabled subgroup analysis to identify possible sources of heterogeneity.

The outcomes undergoing metanalysis were PROMs (adequately reported in 17 studies), maximum phonation time (MPT; adequately reported in 13 studies), and jitter/shimmer (adequately reported in nine studies). We used a random-effects model in all cases, reporting mean difference for all but PROMs, which required reporting the SMD owing to the different scales reported (Voice Handicap Index [VHI]-10, VHI-30, Voice-Related Quality of Life [VRQoL]). Reporting bias was assessed through visual inspection of funnel plots for each meta-analysis. For all other outcomes, a descriptive, narrative synthesis was undertaken.

RESULTS

Study selection

Figure 2 shows the flow of study inclusion and exclusion, as detailed in the PRISMA 2020 statement.[21] Of the 48 full texts retrieved, 23 were included in the present review. Most exclusions were due to not meeting population inclusion criteria; six studies[31–36] included only normophonic subjects, three[37–39] included subjects without a diagnosis of presbyphonia, and two[40,41] included subjects aged < 50 years. Seven papers were excluded because of their design; four[42–45] included only five or fewer participants, and three[46–48] were not interventional. Five papers[49–53] included patients with additional laryngeal or neurological pathology. Finally, two papers[54,55] involved neuromuscular/functional electronic stimulation, which were set out in exclusion criteria as not representing a behavioral

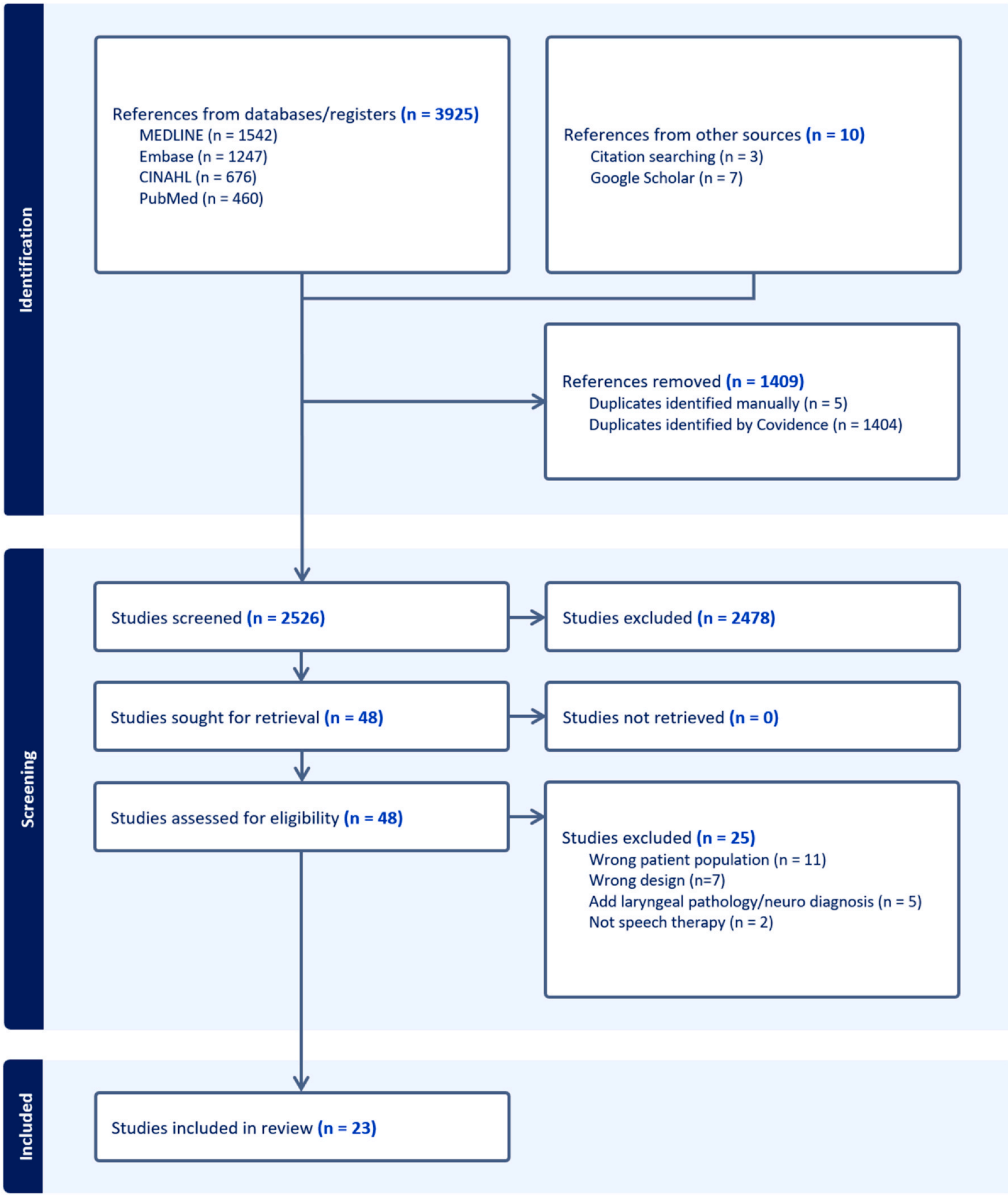


FIGURE 2. PRISMA flowchart for study inclusion.

speech-language pathologist–led intervention. On full text review, B.S.K. and F.G. demonstrated excellent intrarater reliability (Cohen's κ: 0.81).

Study characteristics

Table 2 presents the included studies and their characteristics, including country of origin, publication language, study design, level of evidence (as per the National Health and Medical Research Council's guidelines),[56] and population characteristics (sample size, percentage female, and age distribution).

There were 1050 participants across the studies. Sufficient detail was available for 994 participants approximating a mean age of 72.5 (SD: 8.16) years. All but one study[57] reported proportions of male and female participants. We calculated the total group was approximately 51% female. Per-study female percentages ranged from 0% to 85%, with a median percentage of 48.5% (IQR: 17.75).

Risk of bias in studies

Risk of bias, as identified using the RoB 2 and MINORS tools, is presented in Tables 3 and 4, respectively. For

TABLE 2.
Study Characteristics

Study ID	Language	Country	Design	LOE	Participants	n	% ♀	Age
Belsky et al[58]	English	USA	RCT	II	Dysphonic (VHI-10 ≥11) adults aged ≥55 years with vocal fold atrophy	24	63	G1 (n = 11): 75.2 (8.12) G2 (n = 13): 72.1 (7.91)
Berg et al[59]	English	USA	Retrospective case-control	III-3	Adults aged > 60 years with age-related dysphonia	25	52	G1 (n = 19): 73.0 (6.0) G2 (n = 6): 66.0 (6.0)
Bick et al[60]	English	USA	Retrospective case-control	III-3	Adults aged > 50 years with vocal fold atrophy	64	36	G1 (n = 51): 73.0 [16] G2a (n = 6): 62.5 [14] G2b (n = 8): 63.0 [21]
Desjardins et al[61]	English	USA	RCT	II	Adults aged > 50 years with presbyphonia	12	42	G1: 77.5 (4.66); n = 4 G2: 70.5 (7.23); n = 4 G3: 67.5 (8.74); n = 4
Fujimaki et al[62]	English	Japan	RCT	II	Dysphonic adults aged ≥60 years with glottal insufficiency due to vocal cord atrophy	408	47	G1 (n = 199): 73.9 (7.25) G2 (n = 209): 73.3 (7.68)
Godoy et al[63]	English	Brazil	RCT	II	Adults aged ≥60 years with age-related voice complaints	27	85	G1 (n = 15): 73.9 (6.76) G2 (n = 13): 68.8 (6.04)
Gorman et al[64]	English	USA	Case series	IV	Males aged > 60 years with presbylarynx	19	0	68.4 (60–78)
Guzman et al[65]	English	Chile	RCT	II	Adults aged > 70 years with symptoms of presbyphonia for > 1 year	30	66	73 (70–77)
Inoue et al[66]	Japanese	Japan	Retrospective case series	IV	Adults with age-related vocal fold atrophy	41	59	70.8 (59–85)
Iwaki et al[67,68]	Japanese	Japan	Case series	IV	Dysphonic adults aged > 65 years with vocal fold bowing (if male)	21	38	72.4 (67–80)
Kaneko et al[69]	English	Japan	Retrospective case-control	III-3	Adults aged > 65 years with vocal fold atrophy	22	18	G1 (n = 16): 72.9 (65–81) G2 (n = 6): 74.2 (65–85)
Kaneko et al[70]	English	Japan	Case series	IV	Adults aged > 60 years with vocal fold atrophy	10	20	72 (60–87)
Kim et al[71]	Korean	Korea	Case series	IV	Adults aged > 60 years with vocal fold atrophy	34	47	70.6 (60–82)
Kim et al[57]	Korean	Korea	RCT	II		58	NI	

TABLE 2. (*Continued*)

Study ID	Language	Country	Design	LOE	Participants	*n*	% ♀	Age
					Dysphonic adults aged > 55 years with vocal fold atrophy			G1 ($n = 21$): 64.2 G2 ($n = 20$): 65.6 G3: ($n = 17$): 67.5
Kondo et al[72]	Japanese	Japan	Retrospective case series	IV	Adults aged > 55 years with vocal fold atrophy	21	33	69.7 (55–80)
Mato et al[73]	Japanese	Japan	Case series	IV	Dysphonic adults with vocal fold atrophy	30	50	74.6 (6.47)
Mau et al[74]	English	USA	Retrospective case series	IV	Dysphonic adults aged ≥55 years with vocal fold atrophy	54	51	71 (56–91)
Motohashi et al[75]	English	Japan	Retrospective case series	IV	Adults with vocal fold atrophy	42	50	66.7 (11.1)
Nambu et al[76]	Japanese	Japan	Retrospective case series	IV	Adults with glottal insufficiency due to vocal cord atrophy	19	42	71.9 (50–86)
Santos et al[77]	English	Brazil	RCT	II	Institution-dwelling adults aged > 60 years with age-related voice complaints	42	71	G1 ($n = 21$): 79.0 (7.13) G2 ($n = 21$): 79.0 (10.37)
Sauder et al[78]	English	USA	Case series	IV	Dysphonic adults aged > 65 years with vocal fold atrophy	9	22	76.0 (6.9)
Zhu et al[79]	English	China	Retrospective case series	IV	Dysphonic adults aged > 60 years with presbylarynx or hypoadduction	22	55	66.0 (60–76)
Ziegler et al[80]	English	USA	RCT	II	Dysphonic adults aged > 60 years with presbyphonia	16	50	G1 ($n = 6$): 73.2 (6.6) G2 ($n = 5$): 75.8 (4.0) G3 ($n = 5$): 77.6 (8.4)

Abbreviations: LOE, level of evidence; NI, no information; RCT, randomized controlled trial.
The square brackets represents MEDIANS.

TABLE 3.
Quality Assessment: Cochrane Risk of Bias 2 Tool

Study ID	Randomization Process	Deviations From Intended Interventions	Missing Outcome Data	Measurement of the Outcome	Selection of the Reported Result	Overall Bias
Belsky et al	Low	Low	Low	Low	Low	Low
Desjardins et al	Low	Low	Low	Low	Low	Low
Fujimaki et al	Low	Some Concerns	Low	Low	Low	Some Concerns
Godoy et al	Some Concerns	Low	Low	Low	Low	Some Concerns
Guzman et al	Some Concerns	Low	Low	Low	Low	Some Concerns
Kim et al	Low	Some Concerns	Low	Low	Low	Some Concerns
Santos et al	Low	Some Concerns	Low	Low	Low	Some Concerns
Ziegler et al	Low	Some Concerns	Low	Low	Low	Some Concerns

RCTs, 25% were identified to have a low risk of bias, while the remainder ($n = 6$) were identified to carry "Some Concerns" for overall bias. This arose in most cases ($n = 4$) from the "Deviations from intended interventions" domain and tended to reflect a lack of intention-to-treat analysis in the face of loss to follow-up. In all cases, the reviewers judged this posed little potential for a substantial impact on the result. The overall finding of "Some Concerns" for the remaining two studies arose from the "Randomization process" domain, as neither study documented whether their allocation sequence was concealed until all participants were enrolled. The reviewers did not consider that baseline group differences suggested problematic randomization. Initial agreement between the reviewers for RoB 2 domains was substantial (Cohen's κ: 0.8–1.0) for all domains except for "Deviations from intended interventions" (κ: 0.3). Disagreements between reviewers were resolved by discussion.

The remaining papers ($n = 15$) were assessed with MINORS. There were comparative studies and eligible for a maximum score of 24. The rest were noncomparative and eligible for a maximum score of 16. Scores ranged from 8 (Iwaki et al) to 20 (Berg et al). The comparative studies, as expected, attained higher scores (15–20). The most common concern regarding the nonrandomized studies was a failure to report on prospectively calculated study sizes. A majority ($n = 11$) were inadequate in terms of loss to follow-up, either not reporting it ($n = 4$) or reporting levels higher than 5% ($n = 7$). A majority ($n = 9$) were problematic in terms of blinded assessment of outcomes, either not reporting on blinding ($n = 4$) or otherwise reporting unblinded assessment ($n = 5$). Eight of the studies received an inadequate rating for retrospective data collection. Initial interreviewer agreement for MINORS items was slight for items 3, 5, 6, and 7 (Cohen's κ: 0–0.2), moderate for items 1, 2, and 4 (κ: 0.5), and substantial for items 8–12 (κ: 0.7–0.9). As above, disagreements between reviewers were resolved by discussion.

Our protocol did not specify an a priori level of bias/total MINORS score that would disqualify a study from this review. We felt it was beneficial to utilize bias tool scores to inform discussions between the reviewers over quality concerns and the final inclusion/exclusion decision. Following discussion, no papers were excluded due to risk of bias assessment. As this review aimed to capture a wide representation of presbyphonia voice therapy, we considered it more useful to retain the papers included for data extraction and explore quality/study design in subgroup sensitivity analysis whenever possible. Sensitivity analysis of heterogeneity was also used to explore the confidence and certainty of the evidence for outcomes.

Interventions

Study intervention details, including any comparators/control groups, dosing of home practice, number of sessions with the speech-language pathologist, and duration of treatment/time to follow-up, are presented in Table 5. The

TABLE 4.
Quality Assessment—Methodological Index for Non-Randomized Studies (MINORS)

Study ID	1. Clearly stated aim	2. Consecutive participants	3. Prospective data collection	4. Appropriate endpoints	5. Unbiased assessment	6. Appropriate follow-up period	7. Loss to follow-up <5%	8. Prospective calc. of study size	9. Adequate control group	10. Contemporary groups	11. Baseline equivalence	12. Adequate stat. analysis	TOTAL
Berg et al. (2008)	2	2	1	2	2	2	1	0	2	2	2	2	20
Bick et al. (2021)	2	2	1	2	1	2	1	0	1	2	1	2	17
Kaneko et al. (2015)	2	0	1	2	0	2	0	0	2	2	2	2	15
Gorman et al. (2008)	2	2	2	2	2	2	2	0					14
Inoue et al. (2013)	2	2	2	2	2	2	1	0					13
Iwaki et al. (2017)	2	0	2	1	0	2	1	0					8
Kaneko et al. (2021)	2	1	2	2	2	2	0	0					11
Kim et al. (2010)	2	2	2	2	1	2	0	0					11
Kondo et al. (2022)	2	2	1	2	1	2	0	0					10
Mato et al. (2018)	2	2	2	2	0	2	1	0					11
Mau et al. (2010)	2	2	1	2	1	1	2	1					12
Motohashi et al. (2022)	2	2	1	2	2	2	2	0					13
Nambu et al. (2019)	2	2	1	2	0	2	1	0					10
Sauder et al. (2010)	2	2	2	2	2	2	1	0					13
Zhu et al. (2018)	2	2	1	2	1	2	2	0					12

2: Reported and adequate, **1**: Reported and inadequate, **0**: Not reported

0	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24

Only comparative studies eligible

study durations ranged from a single session to several months. Longitudinal postintervention follow-up was included in only two studies (Godoy et al and Iwaki et al), comprising a 1-month and 12-month follow-up, respectively. The number of therapy sessions received by participants ranged from one session (Guzman et al) to 16 (Godoy et al). Home practice regimes differed by therapy task, requiring practice from one to five times per day from 5 to 7 days per week. Tracking adherence to home practice was mentioned in eight studies[57,58,61,63,64,69,78,80]; however, only Ziegler et al reported the outcome of adherence tracking (~89% compliance). For the remaining seven studies, Belsky et al were unable to report due to missing data, and Godoy et al reported only that no participant was fully compliant with home practice. The remaining five studies mentioned adherence tracking but reported no data.

The most frequently utilized intervention was vocal function exercises (VFE),[81] reported by more than half ($n = 14$) of the included studies.[57,59–61,64,66,67,69,72–74,76,78,80] Seven studies employed semi-occluded vocal tract exercises (SOVTEs) either as a primary or adjunctive therapy approach.[60,63,65,72,74,76,77] Phonatory resistance therapy exercises (PhoRTE) were used in two studies,[58,80] while a qualitatively similar approach ("vocal resistance training") was used by Kaneko et al in their electromyographic (EMG) outcomes study.[70] Other approaches such as resonant voice and abdominal breath work were also used as adjunctive therapy in four[59,60,69,74] and two studies,[60,76] respectively.

As detailed in section Synthesis methods, interventions were summarized and specified as per the RTSS. In all cases, the reviewers confined themselves to what was reported for intervention content in each paper, avoiding speculation based on clinical judgment/experience of delivering presbyphonia therapy. Wherever possible, they used the treatment theory espoused by the authors when specifying their intervention. Table 6 details the intervention component targets and the treatment groups to which they belong (*Organ Functions, Skills & Habits, Representations*). A full articulation of the RTSS treatment component specification for each paper, including therapy ingredients and dosing, is in the Supplemental Information.

TABLE 5.
Study Interventions, Dosing, and Duration

		Home Practice			Clinician Input		
Study	Groups and Interventions	Tasks × Reps	Times Per Day	Days Per Week	Number of Sessions	Duration	Postintervention Follow-Up
Belsky et al	PhoRTE (G1, G2)	4 × 10 1 × 0.5–2 min	1	6	4	4 weeks	None
	EMST (G1)	1 × 5	5	5	4	4 weeks	None
Berg et al	VFE (G1)	4 × 2	2	7	4.1 (mean)	5.1 months (mean)	None
	Resonant voice (G1)	NI					
Bick et al	VFE (G1)	4 × 2	2	7	6–10	"Several months"	None
	AB, resonant voice, SOVT (G1)	NI					
Desjardins et al	VFE (G1)	4 × 4	2	7	4	4 weeks	None
	VFE (G2, G3)	4 × 2	2	7			
	EMST (G2)/IMST (G3)	1 × 5	5	7			
Fujimaki et al	Pulling-phonation (G1)	1 × 2	2	7	3	6 months	None
Godoy et al	Intensive VTE (G1)	5 × 4 min 1 × 2 min	2	7	16	4 weeks	1 month
	Conventional VTE (G2)	5 × 4 min 1 × 2 min	2	7	16	8 weeks	1 month
Gorman et al	VFE	4 × 2	2	7	12	12 weeks	None
Guzman et al	WRT (G1)	6 × 5 min	N/A	N/A	1	N/A	None
Inoue et al	High-load VFE	4 × 3	2	7	NI	4–12 weeks	None
Iwaki et al	VFE (base phase)	4 × 2	2	7	4	8 weeks	1 year
	VFE (maintenance phase)	4 × 2→1	2→1	7→3→1	3–4	7 weeks	
Kaneko et al	VFE (G1)	4 × 2	2	7	NI	~8 weeks	None
	Resonant voice (G1)		NI				
Kaneko et al	Vocal resistance training		NI		5-6	12 weeks	None
Kim et al	SKLCT		NI		4-12 (mean:7.3)	4–12 weeks (mean:7.3)	None
Kim et al	SKLCT (G1)	2 × 1	2	7	8	8 weeks	None
	VFE (G2)	4 × 2	2	7			
Kondo et al	WRT	NI	"daily"	7	6 ($n = 10$)	12–24+ weeks	None
	VFE	4 × 2	"daily"	7	7+ ($n = 11$)		
Mato et al	VFE	4 × 2	2	7	3	8 weeks	None
Mau et al	LSVT/ VFE/resonant voice/WRT		NI		1–7 (mean: 2.6)	NI	None
Motohashi et al	Breath-holding + pulling	1 × 10	3	7	2	4 weeks	None
Nambu et al	VFE	4 × NI	NI	NI	5–12 (mean:8.1)	NI	None
	AB/SOVT/deconstriction		NI				
Santos et al	WRT (G1)		NI		8	8 weeks	None
	Vocal hygiene (G2)		NI				
Sauder et al	VFE	4 × 2	2	7	6	6 weeks	None
Zhu et al	Breath-holding + hard onset	1 × 10 min	2	7	6	12 weeks	None
Ziegler et al	VFE (G1)	4 × 2	2	7	4	6 weeks	None
	PhoRTE (G2)	5 × 2	2	7			

Abbreviations: AB, abdominal breathing; EMST/IMST, expiratory muscle strength training/inspiratory muscle strength training; G1, G2, G3, groups 1, 2, 3, etc; LSVT, Lee Silverman Voice Therapy; NI, no information; PhoRTE, phonatory resistance training exercises; SOVT, semi-occluded vocal tract exercises; VFE, vocal function exercises; WRT, water resistance therapy.

TABLE 6.
RTSS Intervention Targets and Treatment Categories

RTSS Treatment Group →		ORGAN FUNCTIONS					SKILLS & HABITS											REPRESENTATIONS		
Study	RTSS Targets → / ↓ Groups and Interventions (Treatment Group)	↑ glottic closure	↑ muscle strength/activation: larynx	↑ maximum respiratory pressures	↑ loudness	↓ muscle activation: hyperfunction	↑ abdominal movement	↑ respiratory coordination dur. voicing	↓ phonatory irregularities dur. pitch glides	↑ pitch range	↑ loudness/ loudness range	↓ muscle activation levels: larynx	↑ mean airflow during phonation	↑ glottal closure in voicing (hard onset)	↑ accuracy of voice quality discrimination	↑ forward resonance	↑ voice quality (↓ strain/roughness)	↑ adherence to home program	↑ knowledge of prob/ compensation strategies	↑ knowledge of vocal hygiene
Belsky et al. (2021)	G1: PhoRTE + EMST (*OF*)		•	•							•							•		
	G2: PhoRTE (*OF*)		•	•							•							•		
Berg et al. (2008)	G1: VFE/RV (*S&H*)							•	•							•			•	
	G2: No therapy (*C*)																			
Bick et al. (2021)	G1: VFE/AB/SOVTs (*S&H*)						•	•	•				•			•	•		•	•
	G2: No therapy (*C*)																			
Desjardins et al. (2022)	G1: VFE (*S&H*)							•	•									•		
	G2: VFE + EMST (*OF*)			•				•	•									•		
	G3: VFE + IMST (*OF*)			•				•	•									•		
Fujimaki et al. (2017)	G1: Pulling-Phonation (*OF*)		•															•	•	
	G2: Vocal education (*R*)																		•	
Godoy et al. (2019)	G1: Intensive VTE (*OF*)	•	•			•			•	•	•					•	•	•		
	G2: Conventional VTE (*OF*)	•	•			•			•	•	•					•	•	•		
Gorman et al. (2008)	VFE (*S&H*)							•	•									•		
Guzman et al. (2018)	G1: WRT (*OF*)				•											•				
	G2: Open vowels (*C*)																			
Inoue et al. (2013)	High-load VFE (*OF*)		•					•	•											
Iwaki et al. (2017)	VFE (*S&H*)							•	•											
Kaneko et al. (2015)	G1: VFE/RV (*S&H*)							•	•							•		•	•	•
	G2: No therapy (*C*)																			
Kaneko et al. (2021)	Vocal resistance training (*OF*)		•													•			•	
Kim et al. (2010)	SKLCT (*S&H*)						•	•		•	•	•							•	
Kim et al. (2015)	G1: SKLCT (*S&H*)						•	•		•	•	•			•			•	•	
	G2: VFE (*S&H*)							•	•									•		
	G3: No therapy (*C*)																			
Kondo et al. (2022)	WRT/VFE (*S&H*)							•	•			•				•	•			
Mato et al. (2018)	VFE (*S&H*)							•	•										•	•
Mau et al. (2010)	LSVT/ VFE/RV/WRT (*OF*)	•						•	•							•		•		
Motohashi et al. (2022)	Breath-holding + pulling (*OF*)		•																	
Nambu et al. (2019)	VFE/AB/SOVT/Deconstriction (*S&H*)						•	•	•			•					•		•	
Santos et al. (2014)	G1: WRT (*OF*)		•																	
	G2: Vocal education (*R*)																		•	•
Sauder et al. (2010)	VFE (*S&H*)							•	•									•		
Zhu et al. (2018)	Breath-holding + hard onset (*OF*)		•			•	•					•		•					•	
Ziegler et al. (2014)	G1: VFE (*S&H*)							•	•									•		
	G2: PhoRTE (*OF*)		•	•							•							•		
	G3: No Therapy (*C*)																			

Notes: Predominant treatment categories: *OF,* Organ Functions; *S&H,* Skills & Habits; *R*, Representations; *C*, Control (no treatment).
Abbreviations: AB, abdominal breathing; EMST/IMST, expiratory muscle strength training/inspiratory muscle strength training; LSVT, Lee Silverman Voice Treatment; PhoRTE, Phonation Resistance Training Exercises; RTSS, Rehabilitation Treatment Specification System; RV, resonant voice; SKLCT, Seong Kim Laryngeal Calibration Technique; SOVT, semi-occluded vocal tract exercises; VFE, vocal function exercises; VTE, vocal therapy for the elderly; WRT, water resistance therapy.

RTSS-based specification uncovers the overlapping treatment targets of otherwise disparately described therapy approaches. Twenty studies included a *Skills & Habits* treatment target, while only 12 employed an *Organ Functions* target. Eighteen studies explicitly mentioned a treatment component with a *Representations* target.

Skills & Habits targets for the treatments described in the included studies involve volitional action *by* the patient to improve their ability to perform a task, and the mechanism of action for such treatment components is understood to be "learning by doing."[19] Typically, the treatment ingredients for these targets involve an opportunity to repeatedly practice a target behavior in the presence of clinician-supplied instruction, cues, and feedback. Table 6 includes target skills related to respiration (increased abdominal movement, improved phonatory-respiratory coordination), phonation (reduced irregularities in voicing, improved voice quality, increased pitch range, increased loudness/loudness range), resonance (improved forward placement), muscle activation (reduced muscle activation during phonation, improved closure [hard onset]), and perceptual abilities (improved discrimination of voice quality). VFEs, as specified according to RTSS by Van Stan and Stemple (*in review*), require the patient to produce a resonant voice quality, as softly as possible for as long as possible and is therefore an example of a predominantly *Skills & Habits* intervention. The changes that VFE are thought to induce occur through volitional activity in learning a skill by doing.

Organ Functions targets, by contrast, may or may not involve volitional action to effect the intended change, but the change, itself, is not due to learning.[19] The majority of the treatment components with an *Organ Functions* target involve increasing strength or muscular activation through the application of a load of resistance. Examples of predominantly *Organ Functions* interventions are PhoRTE, breath-holding with pulling, hard onset therapy, and respiratory muscle strength training, which seek to reverse laryngeal/respiratory muscle wasting through high-intensity phonation or by directly loading the respiratory and/or adductory musculature.

Finally, *Representation* targets involve a change in knowledge or attitude, as well as choices and decisions that might be linked to an action. This could be the predominant treatment group for an intervention, for example, seeking to reduce vocal misuse (as used in the vocal hygiene control group in Santos et al). The *Representations* domain is also where "volitional targets/ingredients" sit in the RTSS; these are targets that aim at accurate/adherent practice of a treatment. When present, these ingredients might include therapist-delivered feedback on task performance. When the therapist is not present, carrying out the volitional behavior (the therapeutic task) can become a separate target of its own. The RTSS requires that any intervention involving home practice, away from clinician feedback, must include a specified *Representations* target to ensure accurate home practice. Table 6 details *Representations* targets of this type (improved accuracy of home practice, improved adherence to home practice schedule) as well as improved knowledge around the diagnosis of presbyphonia, coping strategies, and vocal hygiene. Every included study likely had a *Representations* target/ingredient, such as online clinician feedback or written instructions for home practice, but only those that made explicit reference to this have been included in Table 6.

The reader will notice that "reducing muscle activation" is listed as both an *Organ Functions* target and a *Skills & Habits* target. This is because the desired change was achieved via different mechanisms of action across studies. Godoy et al and Zhu et al achieved this in a nonvolitional manner; the former used lip trills to achieve "normotensive vocal folds," while the latter utilized acupuncture and neck massage. The reduction in muscle activation was nonvolitional for the patient receiving the therapy; it "just happened" as a direct result of the activity itself, and no training occurred to further improve, generalize, or habituate voice-related changes. In contrast, other studies reduced muscular activation through the volitional practice of a new skill. For example, Kim et al used laughter and inspiratory phonation to decrease hyperfunction and generalize this into speech, and Nambu et al used yawn-sigh in a similar manner.

Investigators also used the same therapy tasks but to different ends (targets) within different treatment groups, meaning the same task represented different therapeutic ingredients. Take, for example, SOVTE. Godoy et al used WRT in 2 cm of water to normalize laryngeal muscle activation (*Organ Functions*). The critical ingredient of the SOVTE was that it provided resistance to phonation. Kaneko et al used the same technique (WRT in 2 cm of water) to improve forward resonance (*Skills & Habits*). The critical ingredient of the SOVTE here was that it provided increased kinesthetic awareness during practice from the tube vibrating and the water bubbling. Both are valid uses of the same technique, but detailing treatment components via the RTSS exemplifies that the same task is being utilized to deliver different therapy ingredients (eg, phonation against resistance versus opportunities to practice voicing with amplified kinesthetic sensations).

As regards categorizing the included interventions into predominant treatment groups, the 14 treatments that employed an *Organ Functions* target were determined by the authors to represent a predominantly *Organ Functions* treatment, while the remaining 14 active treatments were determined to represent predominantly *Skills & Habits* therapies. Although two control group activities were determined to reflect a *Representations* focus, no active treatments belonged to this category (see Table 6).

RESULTS OF INDIVIDUAL STUDIES

Appendices B1-B6 detail effect sizes (Hedge's g) calculated from the results reported in the included studies for

Study or Subgroup	Pre-Therapy Mean	SD	Total	Post-Therapy Mean	SD	Total	Weight	Std. Mean Difference IV, Random, 95% CI
Belsky 2021	21.9363	3.488	16	12.25	3.6528	16	4.2%	2.64 [1.66, 3.62]
Berg 2008	-58	24.1	19	-77	18.9	19	6.7%	0.86 [0.19, 1.53]
Desjardins 2022	19.6667	7.6214	12	16.4167	7.8214	12	5.4%	0.41 [-0.40, 1.22]
Godoy 2019	-82.6976	12.9271	25	-92.0032	11.317	25	7.8%	0.75 [0.18, 1.33]
Inoue 2013	19.39	8.71	41	11.36	7.75	41	9.3%	0.96 [0.51, 1.42]
Kaneko 2015	18.88	8.49	16	7.56	4.79	16	5.4%	1.60 [0.79, 2.41]
Kaneko 2021	17	8.82	10	10.2	3.22	10	4.4%	0.98 [0.04, 1.92]
Kim 2015	29.34	26.5152	41	18.2427	17.4069	41	9.5%	0.49 [0.05, 0.93]
Kondo 2022	19	9.54	21	7.67	5.65	21	6.6%	1.42 [0.73, 2.10]
Mato 2018	25.47	20.22	30	17.33	17.13	30	8.6%	0.43 [-0.08, 0.94]
Motohashi 2022	18.1	7.2	42	12.5	6.3	42	9.4%	0.82 [0.37, 1.27]
Nambu 2019	40.1	25.2	19	20.5	15.8	19	6.7%	0.91 [0.24, 1.58]
Sauder 2010	39.11	21.35	9	23.44	19.54	9	4.3%	0.73 [-0.23, 1.69]
Zhu 2018	22	7.07	22	12.2	6.8	22	6.8%	1.39 [0.72, 2.05]
Ziegler 2014	-84.3	11.6223	11	-90.9091	9.0158	11	5.0%	0.61 [-0.25, 1.47]
Total (95% CI)			**334**			**334**	**100.0%**	**0.94 [0.70, 1.18]**

Heterogeneity: Tau² = 0.11; Chi² = 28.36, df = 14 (P = 0.01); I² = 51%
Test for overall effect: Z = 7.67 (P < 0.00001)

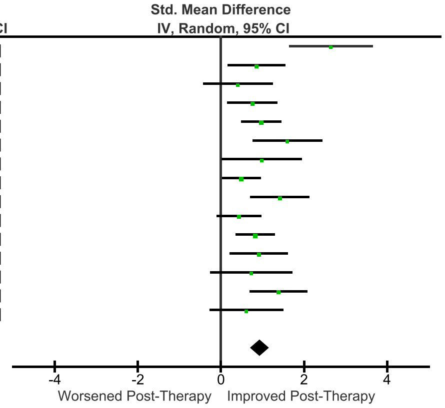


FIGURE 3. PROMs meta-analysis.

PROMs, aerodynamics, acoustics, respiratory measures, laryngeal/glottic function, and subjective/expert-rated measures of voice and functioning. These will be summarized and synthesized in the following sections. For studies that reported multiple time points[66]/longitudinal follow-up,[63,68] we include only preintervention and immediately postintervention outcomes in tabulated results to simplify comparison.

Results of syntheses

Patient-related outcome measures

As demonstrated in Appendix B1, 17 studies reported on a patient-reported outcome measure, 13 using either the VHI-10[82] or VHI-30.[83] All studies reporting on PROMs reported an improvement in at least one group undergoing voice therapy for presbyphonia. Effect sizes ranged from 0.26 (small) to 2.82 (very large), with the majority reporting a large effect in at least one measure and for one subgroup. For those studies including a “no therapy” control group,[57,60,69] no significant change in PROMs were seen.

VHI-10, VHI-30, and VRQoL,[84] reported by 17 studies, were analyzed in a meta-analysis. This was judged appropriate, as VHI-10 has been found to be highly correlated to VHI-30[82] and to VRQoL.[85] As VRQoL scales are in the opposite direction to VHI-10/VHI-30 (higher numbers represent less handicap), VRQoL scores were transformed by subtracting from 0. Also, for the purposes of meta-analysis, comparative cohorts in studies receiving an active behavioral treatment have been combined (unless subgroup sensitivity analysis was being undertaken and the cohorts represented different categories, in which case they were considered separately, ie, sex composition or RTSS treatment group).

The results of the meta-analysis are seen in Figure 3. Analysis confirms a posttherapy SMD for included studies of 0.93 (P < 0.00001; 95% confidence interval [CI]: 0.70–1.17), indicating a large treatment effect. Heterogeneity (I^2) was found to be mildly raised at 51%; it appears that Belsky et al might be an outlier in reporting a very large effect size. When this study is removed, the I^2 drops to 17%, and the results continue to suggest a large treatment effect (SMD: 0.84; P < 0.00001; 95% CI:0.66–1.03). Visual inspection of a funnel plot for this metanalysis (Appendix C1) identified no significant evidence of reporting bias.

Subgroup analysis was carried out according to RTSS treatment group, duration of treatment, number of sessions, sex, and study design. With Belsky et al removed as an outlier, greater treatment effects were reported for studies with predominantly male participants (1.25 versus 0.81; $P = 0.05$) and studies with a longer treatment period (> 12 weeks; SMD 1.18 versus 0.74; $P = 0.03$); however, the same relationship did not hold for number of therapist-led sessions ($P = 0.80$). Study design analysis identified that RCTs were more homogenous in PROMs results and tended to report more modest improvements than non-RCTs (SMD 0.57 versus 0.96; $P = 0.04$). RTSS treatment group was not found to affect PROMs outcomes.

MPT and aerodynamics

The next most reported outcome was MPT, reported by 15 studies (Appendix B2). All but one of these studies reported an improvement in MPT for participants receiving voice therapy, with effect sizes from 0.34 (small) to 1.74 (very large). Although MPT is not, itself, an aerodynamic measure, as it is an indirect correlate of mean laryngeal flow rate and affected by laryngeal airway resistance,[86,87] we include it in this section. As for other aerodynamic outcomes, 11 studies directly reported mean flow rate, with most ($n = 6$) reporting no change. Subglottic pressure was reported by five studies, with three reporting an increase and another, Desjardins, reporting equivocal results. Finally, Guzman et al also reported an increase in glottal resistance following WRT.

A meta-analysis of MPT was undertaken for the 13 studies that reported their results with adequate detail (means/SD). The results confirm a mean increase in MPT of 5.37 seconds (P < 0.00001; 95% CI: 3.52–7.22). Although visual inspection of a funnel plot for these results (Appendix C2) did not reveal evidence of reporting bias, heterogeneity for the included studies was unacceptably high at I^2 = 79%. Gorman et al, the only study whose

Study or Subgroup	Post-Therapy Mean [seconds]	Post-Therapy SD [seconds]	Post-Therapy Total	Pre-Therapy Mean [seconds]	Pre-Therapy SD [seconds]	Pre-Therapy Total	Weight	Mean Difference IV, Random, 95% CI
4.1.1 Organ Functions								
Fujimaki 2017	20.3407	4.724	199	11.9724	5.568	199	11.5%	8.37 [7.35, 9.38]
Inoue 2013	21.48	6.9	41	14.97	6.22	41	9.2%	6.51 [3.67, 9.35]
Kaneko 2021	15.3	5.01	10	10.3	4.03	10	7.5%	5.00 [1.01, 8.99]
Motohashi 2022	20.3	9.3	42	14.2	6.6	42	8.3%	6.10 [2.65, 9.55]
Zhu 2018	17.85	5.53	22	10.35	3.35	22	9.4%	7.50 [4.80, 10.20]
Subtotal (95% CI)			**314**			**314**	**45.9%**	**7.52 [6.39, 8.65]**
Heterogeneity: Tau² = 0.34; Chi² = 4.87, df = 4 (P = 0.30); I² = 18%								
Test for overall effect: Z = 13.04 (P < 0.00001)								
4.1.2 Skills & Habits								
Kaneko 2015	22.25	7.86	16	14.19	5.86	16	6.4%	8.06 [3.26, 12.86]
Kim 2010	17.61	6.07	18	14.79	6.63	18	7.2%	2.82 [-1.33, 6.97]
Kim 2015	15.6163	5.4457	41	13.319	5.9026	41	9.8%	2.30 [-0.16, 4.76]
Kondo 2022	14.84	4.77	21	12.33	5.57	21	8.7%	2.51 [-0.63, 5.65]
Mato 2018	17.48	6.98	30	13.83	6.88	30	8.2%	3.65 [0.14, 7.16]
Nambu 2019	13.2	6.2	19	11.2	5.4	19	7.9%	2.00 [-1.70, 5.70]
Sauder 2010	18.62	5.81	9	17.78	5.35	9	6.0%	0.84 [-4.32, 6.00]
Subtotal (95% CI)			**154**			**154**	**54.1%**	**2.90 [1.56, 4.23]**
Heterogeneity: Tau² = 0.00; Chi² = 5.74, df = 6 (P = 0.45); I² = 0%								
Test for overall effect: Z = 4.25 (P < 0.0001)								
Total (95% CI)			**468**			**468**	**100.0%**	**4.81 [3.04, 6.59]**
Heterogeneity: Tau² = 6.84; Chi² = 47.79, df = 11 (P < 0.00001); I² = 77%								
Test for overall effect: Z = 5.32 (P < 0.00001)								
Test for subgroup differences: Chi² = 26.81, df = 1 (P < 0.00001), I² = 96.3%								

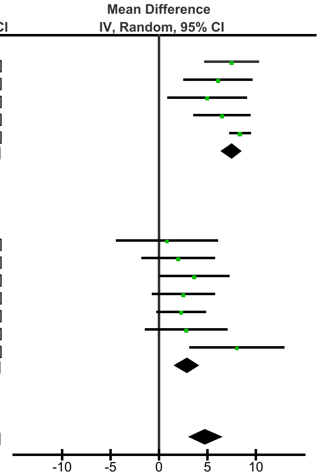


FIGURE 4. RTSS subgroup meta-analysis for maximum phonation time (MPT).

cohort was 100% males, appears to be an outlier. The pretherapy mean MPT for their cohort was 22.0 seconds (SD: 6.86), which is significantly higher than the pretherapy MPT calculated for the participants across the studies, 12.7 seconds (SD: 6.73, $n = 487$). However, removing this study only modestly reduces heterogeneity ($I^2 = 77\%$).

Subgroup analysis proved more telling. With Gorman et al removed, clustering the studies along RTSS treatment group (Figure 4) resulted in significantly different subgroups ($P < 0.00001$) with much lower heterogeneity (*Organ Functions:* $I^2 = 18\%$, *Skills & Habits*: $I^2 = 0\%$). The mean increase to MPT for studies with an *Organ Functions* treatment focus was 7.52 seconds ($P < 0.00001$; 95% CI: 6.39–8.65) compared to an increase of only 2.90 seconds ($P < 0.0001$; 95% CI: 1.56–4.23) for studies with a *Skills & Habits* focus. Analysis per number of sessions was also significant. MPT improved more in studies with *fewer* (< 6) therapist-led sessions (3.26 versus 6.61 seconds; $P = 0.01$). Although this seems counterintuitive, this effect may be in part because all but one *Organ Functions* study featured fewer than six therapy sessions. Subgroup analysis along other categories (duration, sex, and study design) did not yield significant differences.

Acoustic measures

Appendix B3 summarizes the outcomes from the included studies for acoustic outcomes. Measures included shimmer/amplitude perturbation quotient (13 studies), jitter (12 studies), noise-to-harmonics ratio (nine studies), loudness (SPL) or loudness range (seven studies), speaking fundamental frequency (SF_0; seven studies), fundamental frequency (f_0) range (six studies), and cepstral measures (two studies). Of the 16 studies reporting acoustic outcomes, only two failed to identify an improvement posttherapy, although effect sizes varied considerably from small (0.32) to extremely large (3.34).

The measures with the least convincing pattern of change were SF_0 and loudness. Seven studies investigated SF_0, and only two reported a change. Both Bick et al and Zhu et al reported an increase in pitch (females only in the former case). In the six studies reporting SPL, only two demonstrated an improvement (increase). SPL and F_o range, however, showed more convincing change. F_0 range improved in five of the six studies that investigated it. SPL range, investigated in one study, improved with small-medium effect size.

The outcomes for jitter and shimmer were synthesized in a meta-analysis. Figure 5A demonstrates that jitter decreased by a mean of 0.62% ($P = 0.0007$; 95% CI: 0.26%–0.97%), but heterogeneity was high with $I^2 = 80\%$. The results for shimmer (Figure 5B) were more consistent across the studies, with a mean decrease of 1.05% ($P < 0.00001$; 95% CI: 0.67%–1.44%) and an I^2 of 27%. Inspection of funnel plots (Appendix C3 and 4) revealed a qualitatively less symmetrical distribution of results and leaves open the possibility of publication bias. Acoustic measure subgroup analysis, perhaps hampered by fewer eligible studies, did not prove to be significant in any category.

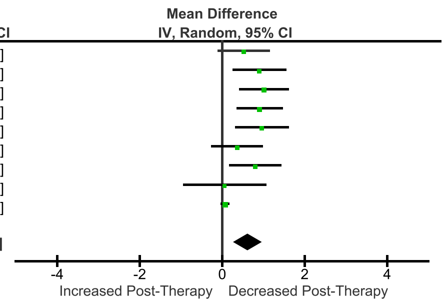
Expert-rated outcomes, glottic/laryngeal behavior, and respiratory measures

Appendices B4-B6 summarize the outcomes and effect sizes for expert-rated perceptual voice outcomes, glottic/laryngeal behavior, and respiratory measures, respectively.

Twelve studies reported on a measure of perceptual voice quality (Appendix B4), with posttherapy improvement noted in at least one auditory-perceptual domain for at least one group receiving active voice therapy treatment for all studies. Most studies ($n = 9$) used the GRBAS (grade, roughness, breathiness, asthenia, strain) rating scale (0–3). Two used a visual analog scale (VAS; 0–100) for overall severity or perceptual subcategories. One study used the categories “improved/worsened/the same” listening to paired recordings. Owing to the mix of categorical, continuous (VAS), and ordinal data (despite many studies treating the GRBAS scale as continuous), we did not feel

Study or Subgroup	Pre-Therapy Mean [%]	SD [%]	Total	Post-Therapy Mean [%]	SD [%]	Total	Weight	Mean Difference IV, Random, 95% CI
Kaneko 2015	1.41	1	16	0.89	0.72	16	10.9%	0.52 [-0.08, 1.12]
Kaneko 2021	1.34	0.98	10	0.44	0.23	10	10.7%	0.90 [0.28, 1.52]
Kim 2010	2.1306	1.4973	34	1.1188	0.8339	34	11.2%	1.01 [0.44, 1.59]
Kim 2015	2.121	1.4647	41	1.2129	0.9336	41	11.7%	0.91 [0.38, 1.44]
Kondo 2022	2.19	1.33	21	1.227	0.6	21	10.7%	0.96 [0.34, 1.59]
Mato 2018	1.91	1.26	30	1.55	1.11	30	10.9%	0.36 [-0.24, 0.96]
Motohashi 2022	2.5	1.6	42	1.7	1.2	42	10.9%	0.80 [0.20, 1.40]
Sauder 2010	1.28	1.23	9	1.22	0.862	9	7.2%	0.06 [-0.92, 1.04]
Zhu 2018	0.56	0.08	22	0.49	0.16	22	15.8%	0.07 [-0.00, 0.14]
Total (95% CI)			**225**			**225**	**100.0%**	**0.62 [0.26, 0.97]**

Heterogeneity: Tau² = 0.21; Chi² = 39.20, df = 8 (P < 0.00001); I² = 80%
Test for overall effect: Z = 3.41 (P = 0.0007)



Study or Subgroup	Pre-Therapy Mean [%]	SD [%]	Total	Post-Therapy Mean [%]	SD [%]	Total	Weight	Mean Difference IV, Random, 95% CI
Desjardins 2022	4.2467	1.8326	12	3.4033	1.6047	12	6.6%	0.84 [-0.53, 2.22]
Kaneko 2015	3.8	2.4	16	3.01	2.22	16	5.1%	0.79 [-0.81, 2.39]
Kaneko 2021	5	2.34	10	2.7	1.6	10	4.3%	2.30 [0.54, 4.06]
Kim 2010	5.4694	3.0186	34	3.3406	0.9637	34	9.9%	2.13 [1.06, 3.19]
Kim 2015	5.3312	4.1711	41	3.7778	2.1728	41	6.1%	1.55 [0.11, 2.99]
Kondo 2022	4.13	3.86	21	3.15	2.48	21	3.5%	0.98 [-0.98, 2.94]
Mato 2018	4.67	1.82	30	3.71	1.31	30	14.9%	0.96 [0.16, 1.76]
Motohashi 2022	5.9	3.5	42	4.8	3.3	42	6.0%	1.10 [-0.35, 2.55]
Sauder 2010	4.98	2.27	9	3.84	2.41	9	2.9%	1.14 [-1.02, 3.30]
Zhu 2018	2.07	0.24	22	1.39	0.15	22	40.8%	0.68 [0.56, 0.80]
Total (95% CI)			**237**			**237**	**100.0%**	**1.05 [0.67, 1.44]**

Heterogeneity: Tau² = 0.09; Chi² = 12.41, df = 9 (P = 0.19); I² = 27%
Test for overall effect: Z = 5.38 (P < 0.00001)

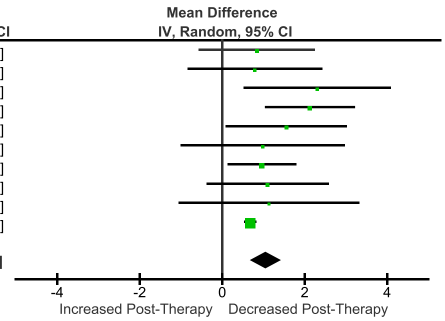


FIGURE 5. **A** (top) and **B** (bottom): Meta-analysis for Jitter (top) and Shimmer (bottom).

such data were appropriate for meta-analysis and reported these findings in narrative synthesis.

Dysphonia severity (variously reported as "overall severity" from CAPE-V, "G (*grade*)" from GRBAS, or the "GRBAS total" sum of subscales) was the most reported category (12 studies). All but one paper reported an improvement following voice therapy with large to extremely large effect sizes. Breathiness was the next most reported metric (seven studies), with six reporting improvement (reduction). Roughness, asthenia, and strain were reported in six studies; roughness improved in all six studies, asthenia improved in four, and strain improved in three. Finally, improvement was also found in vocal instability (two studies) and perceptual loudness (one study). Although not a perceptual voice assessment, Mau et al reported on expert (ie, clinician) rating of patient functioning and participation using the National Outcome Measures Scales, finding significant ($P < 0.001$) improvement posttherapy.

Ten papers reported glottal characteristics/laryngeal function (Appendix B5), although methods and measures varied widely. Metrics included electroglottography (CQ),[65] laryngeal EMG,[70] measures derived from videostrobosopy,[57,61,63,69,71,73,78] and aerodynamic assessment.[64] Videostroboscopic measures included structured perceptual assessment tools, such as the Voice-Vibratory Assessment with Laryngeal Imaging,[61] semistructured and ad-hoc expert-rated scales,[57,63,71,78] as well as calculated parameters such as vibratory dynamics,[78] bowing index (a measure of the degree of vocal fold bowing at rest),[88] normalized mucosal wave amplitude, and normalized glottic gap.[70]

The most reported glottal behavior outcome was glottic closure. Eight studies reported this, including six that reported increased closure after therapy. The results for vocal fold bowing were more equivocal—fewer studies reported on it (four), and only half of these found a posttherapy improvement. Godoy et al only found such an improvement in the intensive therapy group (four sessions instead of two sessions per week). Kaneko et al identified increased muscle activation (via laryngeal EMG) and improved vibratory characteristics following high-intensity "vocal resistance training" despite no change to bowing. Three other studies investigated mucosal wave characteristics, and two of these also identified posttherapy improvements in amplitude, symmetry, or regularity. Compensatory hyperfunction was assessed in five studies and found to improve in three.

Finally, Appendix B6 reports on respiratory measures, the least-reported outcome category. Two papers investigated respiratory muscle strength, Belsky et al and Desjardins et al, as both studies investigated the impact of this form of strength training on presbyphonia. Those interventions targeting improved expiratory muscle strength resulted in greater improvement to maximum expiratory pressure. Those that targeted inspiratory muscle strength resulted in larger improvements to maximum inspiratory pressure, in line with current evidence for respiratory muscle strength training interventions.[89] Three papers investigated pulmonary function, but only one, Santos et al, identified an improvement to vital capacity following voice therapy. No papers reported on speech-breathing kinematics.

DISCUSSION

This paper reviewed voice therapy literature to understand how clinician-delivered behavioral interventions affect outcomes in presbyphonia. A detailed systematic search was undertaken, and the relevant studies appraised by two reviewers. The resultant 23 studies, the largest body of presbyphonia voice therapy research synthesized to date, represented a variety of designs and a range of quality. Rather than exclude nonrandomized studies, or disregard papers of comparatively lower quality, we decided to explore issues of heterogeneity and potential bias through holistic appraisal and sensitivity analysis. To facilitate meaningful comparison of the therapeutic approaches, a consensus-based specification methodology was implemented, applying the RTSS, to understand the therapeutic targets and ingredients involved. This detailed specification enabled theory-based subgrouping alongside analysis based on intervention duration, number of sessions, participant demographics, and study design.

PROMs were the most reported, and perhaps most consistently improved, set of outcomes following voice therapy for presbyphonia. Improving perceived vocal handicap, or VRQoL, is an important therapy goal in the treatment of presbyphonia and may be a more sensitive indicator of real-life improvement in activity and participation.[78] Difficulty communicating for older people can have severe implications on health and social integration,[90] and more than half with a diagnosis of presbyphonia have a significantly impaired quality of life.[1] All studies reporting on PROMs in the current review reported improvement; for those studies comparing an active treatment and no-treatment control group, significant improvement was only identified in the therapy group. However, as only four of the 17 studies reporting on PROMs included a no-therapy control group, we must interpret these results cautiously. It is uncertain how much the improvements were attributable to active therapeutic ingredients and how much were due to observer-expectancy or placebo effects, to which voice PROMs may be particularly susceptible.[91]

MPT was the next most reported outcome and another measure that improved following voice therapy for presbyphonia. MPT is frequently used to assess vocal function and phonatory mechanics, as it requires minimal specialized equipment. The maximum time a person can sustain phonation is directly impacted by flow rate and the valving capability of the vocal folds.[86,87,92] Although MPT in younger populations has not been shown to be correlated with lung function derived from spirometry, it has been correlated with forced vital capacity (which is related to the strength of respiratory musculature, the elasticity of the lungs and the rigidity of the ribcage) in older adults requiring long-term care.[93] As these factors, laryngeal valving and vital capacity, are affected by the aging process and implicated in the pathogenesis of presbyphonia,[9] MPT is a frequently used metric to measure voice therapy progress in this patient group (with the caveat that an improvement in MPT does not differentiate which aspect, respiratory, laryngeal, or both, has improved). Metanalysis confirmed that for the included studies, voice therapy increased MPT for people with presbyphonia by a mean of 5.37 seconds. When this increase is considered against the cross-study participants' pretreatment mean MPT of 12.7 seconds (particularly as recent data suggest normative MPTs for adults in the seventh to ninth decades is ~22 seconds),[94] it suggests a significant improvement in the direction of normality. As recent work considers a posttherapy increase of greater than 1.41 seconds to be the threshold for treatment success,[95] it is also likely a clinically meaningful one. Caution is required, however, as increased MPT may not necessarily be associated with improved functional outcomes. Physiological improvement still requires generalization of these gains to everyday voice use. Furthermore, the use of MPT has been found to demonstrate considerable intrasubject variability across trials.[86] Although adults with the most severe presbyphonic symptoms also have respiratory impairment,[96,97] studies investigating pulmonary correlates (outside of MPT) were too few to conclude the impact of voice therapy on respiratory function.

Voice quality, reflected in both acoustic and expert-rated auditory-perceptual ratings, improved across the studies that reported these measures. Aging is known to affect perceptual voice quality (reflected in increased strain and/or breathiness)[10] as well as the integrity of the acoustic signal (in terms of increased perturbation measures and the ratio of noise to harmonics).[98] As people with presbyphonia report anxiety and frustration around voice quality changes and the need to repeat themselves,[2] voice quality improvement is a valid goal for presbyphonia voice therapy. Meta-analysis confirmed significant reductions to both jitter (0.62%) and shimmer (1.05%) for participants included in this review, two measures found to differ significantly between older adults with presbyphonia versus younger people.[98] Furthermore, auditory-perceptual outcomes demonstrated consistent improvement in listeners' perceptions of voice quality following voice therapy for presbyphonia, meaning evidence for voice quality improvements was not confined to acoustic measures. Those studies that included assessment of laryngeal appearance and glottic function support that voice quality improvement may arise from improvements to laryngeal *function* (closure pattern, vibratory characteristics, reduced hyperfunction, increased muscle activation) rather than to changes in laryngeal *structure* (reduced vocal fold bowing). Although evidence is comparatively sparse, authors suggest that while voice therapy appears capable of altering respiratory coordination and laryngeal muscle activity, leading to improved function and neuromuscular adaptation,[99] it is perhaps unable to alter the atrophic morphology of aged vocal fold mucosa.[61,70]

Investigating heterogeneity was accomplished through subgroup sensitivity analysis wherever possible. Factors found to significantly contribute to heterogeneity were study design (RCT versus non-RCT), RTSS treatment group (*Organ Functions* versus *Skills & Habits*), treatment duration (< 12 weeks versus ≥12 weeks), and sex of the participants (predominantly male versus mixed). As mentioned, the thresholds for duration, number of sessions, and sex composition were pragmatically selected by the reviewers as the most productive of adequately sized subgroups for comparisons. Subgroup analysis in terms of *actual* home practice dose was not possible, as only one study reported on participant adherence to the prescribed therapy regimen. Analysis by treatment duration (as a proxy for home practice dose) and session number was difficult due to sparse descriptions of these elements, or the inclusion of multiple durations/session numbers due to retrospective study designs. Our subgroup analysis would best be interpreted as indicative, rather than conclusive, evidence for the reported factors.

There is a risk to internal validity with nonrandomized or noncomparative studies. Subgroup analysis based on study design (therefore, quality) identified a significant difference between RCTs and non-RCTs only in terms of PROMs. The included RCTs reported more modest improvements to VHI-10, VHI-30, and VRQoL than the studies with nonrandomized or noncomparative designs. This difference (SMD 0.57 versus 0.96) was significant but only of degree (medium versus large effect size) rather than direction of change. Subgroup analysis of study design did not identify systematic differences in reported outcomes in the remaining meta-analyses (MPT or jitter/shimmer).

Subgroup analysis also identified that RTSS treatment group was associated with significant differences to MPT improvement. Interventions that used an *Organ Functions* approach reported significantly greater gains to MPT than those using a predominantly *Skills & Habits* approach (7.52 seconds vs 2.90 seconds). *Organ Functions* treatments aimed to improve vocal function by taxing or loading the phonatory or respiratory musculature: Kaneko et al employed “vocal resistance therapy,” Inoue et al used high-intensity VFEs, Zhu et al used breath-holding and hard-onsets, Fujimaki et al used pulling forcefully while phonating, and Motohashi et al used breath-holding and pulling (see Supplemental Information for detail). Studies with *Skills & Habits* interventions all employed VFEs (and in the case of Kim et al. (2015), an additional *Skills & Habits* intervention, SKLCT). VFEs include as an ingredient a patient-specific amount of phonation related to vital capacity divided by 80–100 ml/sec of airflow,[100] which has some task similarity with MPT elicitation. Even though all the *Skills & Habits* treatments employed a sustained phonation ingredient, the *Organ Functions* studies (three of which did not include sustained phonation) appeared superior at improving MPT. This is even more remarkable considering Motohashiet al's approach not only did not employ an MPT ingredient, it did not involve phonation! Since MPT is affected by laryngeal and respiratory pathogenic factors thought to contribute to presbyphonia, this result tentatively suggests that therapeutic approaches with an *Organ Functions* target, including a dose of resistance/maximal adduction in their ingredients, may address the physiological basis of presbyphonia better than approaches with a *Skills & Habits* focus. This might operate through improved respiratory function or improved laryngeal closure and valving, or both.

Subgroup analysis also suggested that longer interventions were associated with greater improvements to reported vocal handicap (SMD 1.18 versus 0.74). The same was not true for interventions with more therapist input ($P = 0.80$). It is possible that longer interventions, rather than more therapy sessions, enabled more home practice and led to greater improvements in self-reported measures of vocal handicap. Despite the lack of substantial adherence data, two of the retrospective studies (Berg et al and Bick et al) reported a positive association between improvement and adherence to therapy. Inoue et al and Iwaki et al also reported a trend for continued improvement during longer periods of treatment for not only self-reported VHI but also MPT (despite our nonsignificant MPT duration subgroup analysis). The relationship between improvement in PROMs and adherence to home practice and therapy attendance is likely to be a complicated one, however, and there is also significant potential for placebo and self-bias in such measures.[59,91] There is also the possibility of an observer-expectancy effect, whereby patients engaged in a long therapy process may be unintentionally influenced to report continued gains. This was precisely what Iwaki et al, reporting in their 2017 study, concluded; despite most physiologically based metrics returning to pretherapy levels at 12 months posttherapy, all their subjects reported sustained improvement to VHI scores.

Subgroup analysis also identified that studies with predominantly male subjects (> 65%) reported larger improvements in PROMs than studies with fewer males. On a per-study level, Kim et al found that males experienced greater improvements in measures of acoustic perturbations, and Mato et al found a near-significant ($P = 0.06$) association between improvement from therapy and being male. This tentative finding is not surprising when considering the literature on aging, respiration, and the voice. Despite males experiencing greater age-related atrophic change to the vocal folds,[8] greater loss of lung elasticity,[101] and a faster rate of respiratory muscle strength decline,[102] they retain larger raw respiratory values and greater strength. Females are 30% more likely than age-matched males to experience presbyphonia from presbylarynx and to present with shorter MPTs.[24] All things being equal, males may tap into greater physiological reserve during therapy and experience marginally greater reduction in handicap from presbyphonia due to sex-based differences in physiology and aging. Males, due to their comparatively larger impairment, may also have more gains to make, whereas females may reach a ceiling effect more quickly.

Study limitations

Although our purposefully broad study design and language of publication inclusion criteria enabled wide access to reports of therapy in the literature, this and our use of exploratory metanalysis were also limitations. What was gained in breadth came at the cost of certainty and internal validity. Many of the included cohorts were also small, and nearly half of the studies were noncomparative in design. Subgroup analysis investigating participant demographic categories other than sex was not possible. Age was the other well-reported demographic, but it was too homogenous to stratify into useful comparative study subgroups. A handful of studies reported demographics, such as employment status,[72,76] comorbidities,[73,74] and duration of illness,[72,73] but these were too few to enable analysis according to these demographic categories. Another consequence of our broad inclusion criteria was that nearly one-third of the included papers were not written in the English language and were translated using machine translation. The results were consistent and comprehensible, but potentially the subtlety of detail was sacrificed in translation.

Another limitation is that among the acoustic measures of voice quality identified in the included studies, only jitter and shimmer were reported by enough studies to synthesize in metanalysis. These time-based measures of perturbation are far less robust and poorly correlated to auditory-perceptual ratings of voice quality than newer metrics such as cepstral peak prominence (CPP).[103] This is why CPP, rather than jitter or shimmer, has been recommended by an American Speech-Language-Hearing Association (ASHA) expert panel for instrumental voice assessment.[104] As mentioned, the funnel plots hinted at possible publication bias in these metrics for the included studies. Although, when considered together with auditory-perceptual outcomes, there remains a pattern of improvement to voice quality following speech-language pathologist–led voice therapy for presbyphonia, this requires confirmation when data using more robust metrics are available in sufficient number.

CONCLUSIONS

This systematic review sought to broadly evaluate the literature on presbyphonia voice therapy. Outcomes included PROMs, aerodynamics, acoustics, laryngeal/glottic function, auditory-perceptual voice quality, and respiratory measures. Notwithstanding varying levels of quality, differing study designs, and the limitations mentioned above, meta-analysis and narrative synthesis demonstrated that speech-language pathologist–delivered behavioral interventions were successful at improving patient perceptions of handicap (PROMs), physiological function (MPT), and voice quality (acoustic perturbations and auditory-perceptual measures). Factors that appeared to affect voice outcomes were RTSS treatment group (*Organ Functions* > *Skills & Habits*), duration of input (greater than 12 weeks > less than 12 weeks), study design (RCT < non-RCT), and the sex composition of the cohort (predominantly male > mixed); however, these findings must be interpreted cautiously due to the methodological limitations of the studies and their small sample sizes.

Although the present review is not the first to systematically interrogate the literature on presbyphonia and its treatment, it is the first to focus exclusively on the effect of speech-language pathologist–led voice therapy and includes the largest body of studies on the topic. Furthermore, employing the RTSS as a framework to describe the included interventions enabled us to tease apart the question of how such therapy works and on what aspects of vocal functioning through subgroup analysis. To the authors' knowledge, the present review is the first time the RTSS has been used in a systematic review of voice therapy; it proved quite useful in providing a rationale for combining approaches with the same treatment group, uncovering otherwise possibly undetected differences in outcomes.

OTHER INFORMATION

Registration and protocol

This study and its protocol are registered on PROSPERO (CRD42023403726).

Support

B.S.K. holds a Chief Nurse Research Fellowship from the Centre for Nursing, Midwifery and Allied Health Professional Research, University College London Hospitals NHS Foundation Trust, which made this study possible. J.V.S.'s work was supported by the National Institutes of Health (NIH) National Institute on Deafness and Other Communication Disorders via Grant R01 DC020247 (principal investigator: Dr. Jarrad Van Stan). The article's contents are solely the responsibility of the authors and do not necessarily represent the official views of the NIH.

Declaration of Competing Interest

B.S.K. confirms that the authors of the above manuscript have no declarations of interest to disclose.

Appendix A

MEDLINE Search Strategy	
1.	Aging/ or aged/ or "aged, 80 and over"/ or frail elderly/ or middle aged/ or Retirement/ or atrophy/ or muscular atrophy/ or Age Factors/
2.	larynx/ or glottis/ or laryngeal cartilages/ or laryngeal mucosa/ or laryngeal muscles/ or voice disorders/ or aphonia/ or dysphonia/ or hoarseness/ or Voice/ or Voice Quality/ or Vocal Cords/ or Phonation/ or Laryngeal Diseases/
3.	(elderly or geriatric or senescent or age-related or "older adult" or "senior citizen" or "old age").tw.
4.	(dysphon* or "voice disord*" or voice or phonat* or hoarse*).tw.

MEDLINE Search Strategy	
5.	(presbyphon* or presbylary* or "vocal fold atrophy" or "vocal fold bowing" or "bowed vocal fold*" or "ag* larynx" or “ag* voice” or “glott* insufficiency” or “glott* gap”).tw.
6.	speech therapy/ or voice training/ or Speech-Language Pathology/
7.	(voice therap* or (speech adj1 therap*) or (speech adj1 path*) or voice train* or vocal* train* or (voice adj1 exercis*) or (vocal adj1 exercis*)).tw.
8.	1 and 2
9.	3 and 4
10.	1 and 4
11.	2 and 3
12.	5 or 8
13.	5 or 9
14.	5 or 10
15.	5 or 11
16.	6 or 7
17.	12 and 16
18.	13 and 16
19.	14 and 16
20.	15 and 16
21.	17 or 18 or 19 or 20
Embase/Emcare search strategy	
1.	Aged/ or atrophy/ or aging/ or frail elderly/ or frailty/ or atrophy/ or muscle atrophy/
2.	Phonation/ or voice/ or vocal cord/ or voice disorder/ or larynx/ or larynx disorder/ or larynx cartilage/ or larynx mucosa/ or larynx epithelium/ or voice analysis/ or glottis/
3.	(elderly or geriatric or senescent or age-related or "older adult" or "senior citizen" or "old age").tw.
4.	(dysphon* or "voice disord*" or voice or phonat* or hoarse*).tw.
5.	(presbyphon* or presbylary* or "vocal fold atrophy" or "vocal fold bowing" or "bowed vocal fold*" or "ag* larynx" or “ag* voice” or “glott* insufficiency” or “glott* gap”).tw
6.	Voice training/ or speech therapy/ or speech rehabilitation/
7.	(voice therap* or (speech adj1 therap*) or (speech adj1 path*) or voice train* or vocal* train* or (voice adj1 exercis*) or (vocal adj1 exercis*)).tw.
8.	1 and 2
9.	3 and 4
10.	1 and 4
11.	2 and 3
12.	5 or 8
13.	5 or 9
14.	5 or 10
15.	5 or 11
16.	6 or 7
17.	12 and 16
18.	13 and 16
19.	14 and 16
20.	15 and 16
21.	17 or 18 or 19 or 20
CINAHL search strategy	
1.	(MH "Aged") OR (MH "Aged, 80 and Over") OR (MH "Health Services for Older Persons") OR (MH "Aging") OR (MH "Middle Age") OR (MH "Rehabilitation, Geriatric") OR (MH "Frail Elderly") OR (MH "Retirement") OR (MH "Atrophy") OR (MH "Muscular Atrophy") OR (MH "Age Factors")
2.	(MH "Larynx") OR (MH "Laryngeal Muscles") OR (MH "Laryngeal Mucosa") OR (MH "Glottis") OR (MH “Laryngeal Cartilages”) OR (MH "Voice") OR (MH "Voice Quality") OR (MH "Voice Disorders") OR (MH "Dysphonia, Muscle Tension") OR (MH "Aphonia") OR (MH "Phonation") OR (MH “Hoarseness”) OR (MH “Laryngeal Diseases”) OR (MH “Vocal Cords”)
3.	TI (elderly or geriatric or senescent or age-related or "older adult" or "senior citizen" or "old age") OR AB (elderly or geriatric or senescent or age-related or "older adult" or "senior citizen" or "old age")
4.	TI (dysphon* or "voice disord*" or voice or phonat* or hoarse*) OR AB (dysphon* or "voice disord*" or voice or phonat* or hoarse*)
5.	(MH “Presbyphonia”)
6.	TI (presbyphon* or presbylary* or "vocal fold atrophy" or "vocal fold bowing" or "bowed vocal fold*" or "ag* larynx" or “ag* voice” or “glott* insufficiency” or “glott* gap”) OR AB (presbyphon* or presbylary* or "vocal fold atrophy" or "vocal fold bowing" or "bowed vocal fold*" or "ag* larynx" or “ag* voice” or “glott* insufficiency” or “glott* gap”)
7.	(MH "Speech-Language Pathology") OR (MH "Rehabilitation, Speech and Language") OR (MH "Speech Therapy") OR (MH “Voice Therapy”)
8.	TI (voice therap* or (speech N1 therap*) or (speech N1 path*) or voice train* or vocal* train* or (voice N1 exercis*) or (vocal N1 exercis*)) OR AB (voice therap* or (speech N1 therap*) or (speech N1 path*) or voice train* or vocal* train* or (voice N1 exercis*) or (vocal N1 exercis*))

MEDLINE Search Strategy	
9.	1 and 2
10.	3 and 4
11.	1 and 4
12.	2 and 3
13.	5 or 6
14.	9 or 13
15.	10 or 13
16.	11 or 13
17.	12 or 13
18.	7 or 8
19.	14 and 18
20.	15 and 18
21.	16 and 18
22.	17 and 18
23	19 or 20 or 21 or 22
Google Scholar	
"presbyphonia OR presbylarynx OR presbylaryngeus OR presbylaryngis OR "vocal fold atrophy" OR "vocal fold bowing" OR "age related voice" OR "age related dysphonia" AND "voice therapy""	

Appendix B1: Patient-Reported Outcome Measure Effect Sizes

Study	Interventions	VHI-10 VHI-30	VRQoL	Others
Belsky (2021)	G1: PhoRTE + EMST	↓ 2.82		↓^{AVI} 2.08
	G2: PhoRTE	↓ 2.72		↓^{AVI} 2.76
Berg (2008)	G1: VFEs/RV		↑ 0.88	
	G2: No therapy		NS	
Bick (2021)	G1: VFEs/AB/SOVTs		↑*	↑***
	G2: No therapy			
Desjardins (2022)	G1: VFEs	NC		↓^{GFI} 0.45‡
	G2: VFEs + EMST	NC		
	G3: VFEs + IMST	↓ 1.41‡		↓^{GFI} 0.66‡ ↑^{CPIB} 2.39‡
Godoy (2019)	G1: Intensive VTE		↑ 1.15	
	G2: Conventional VTE		↑ 0.26	
Inoue (2013)	High-load VFEs	↓ 0.99		
Iwaki (2017)	VFEs	↓*		
Kaneko (2015)	G1: VFEs/Resonant voice	↓ 1.60		
	G2: No therapy	NC		
Kaneko (2021)	Vocal resistance training	↓ 0.98		
Kim (2015)	G1: SKLCT	NC		
	G2: VFEs	↓ 0.48		
	G3: No therapy			
Kondo (2022)	WRT/VFEs	↓ 1.42		
Mato (2018)	VFEs	↓ 0.43		
Motohashi (2022)	Breath-holding + pulling	↓ 0.82		
Nambu (2019)	VFEs/AB/SOVT/deconstriction	↓ 0.91		
Sauder (2010)†	VFEs	↓ 0.73		↓^{PPE} 0.96
Zhu (2018)	Breath-holding + hard onset	↓ 0.87^{W}		
Ziegler (2014)†	G1: VFEs		↑ 0.55	
	G2: PhoRTE		↑ 0.69	↓^{PPE} 0.85
	G3: No Therapy		NC	

† Alpha < 10. ‡ *P not reported.* *$P < 0.05$, ***$P < 0.001$.
Abbreviations: AVI, Aging Voice Index; CPIB, Communicative Participation Item Bank; GFI, Glottal Function Index; NC, no significant change; PPE, Perceived phonatory effort; W, Wilcoxon signed rank effect size.

Appendix B2: MPT and Aerodynamic Outcome Measure Effect Sizes

Study	Interventions	MPT	Flow	Subglottic Pressure	Glottal Resistance
Belsky (2021)	G1: PhoRTE + EMST		NC		
	G2: PhoRTE		NC		
Bick (2021)	G1: VFEs/AB/SOVTs	↑***	NC		
	G2: No therapy				
Desjardins (2022)	G1: VFEs		↑ 0.43‡	↑ 0.89‡	
	G2: VFEs + EMST		NC	NC	
	G3: VFEs + IMST		↓ 0.31‡	↓ 0.70‡	
Fujimaki (2017)	G1: Pulling-phonation	↑ 1.62			
	G2: Vocal education	NC			
Gorman (2008)	VFEs	↑ 1.74	$\downarrow^{CMF}$ 0.21	$\uparrow^{CMF}$ 0.53	
Guzman (2018)	G1: WRT		↑**	↑***	↑***
	G2: Open vowels (control)		NC	NC	NC
Inoue (2013)	High-load VFEs	↑ 0.98			
Iwaki (2017)	VFEs	↑*	↓*		
Kaneko (2015)	G1: VFEs/resonant voice	↑ 1.13	NC		
	G2: No therapy	NC	NC		
Kaneko (2021)	Vocal resistance training	↑1.05	NC	↑ 0.97	
Kim (2010)	SKLCT	↑ ♂:0.43 ↑ ♀:0.99			
Kim (2015)	G1: SKLCT	↑ 0.74	NC	NC	
	G2: VFEs	NC	NC	NC	
	G3: No therapy				
Kondo (2022)	WRT/VFEs	↑ 0.47	NC		
Mato (2018)	VFEs	↑ 0.52			
Motohashi (2022)	Breath-holding + pulling	↑ 0.75	↓ 0.48		
Nambu (2019)	VFEs/AB/SOVT/deconstriction	↑ 0.34			
Sauder (2010)†	VFEs	NC			
Zhu (2018)	Breath-holding + hard onset	↑ 1.61			

† Alpha < 10. ‡ *P* not reported. **P* < 0.05, ***P* < 0.01, ****P* < 0.001. ♂: males. ♀: females.
Abbreviations: CMF, comfortable loudness.

Appendix B3: Acoustic Outcome Measure Effect Sizes

Study	Interventions	Jitter	Shimmer	NHR	CPP/ Cepstral	SPL	SPL Range	Speaking F_0	F_0 Range
Belsky (2021)	G1: PhoRTE + EMST				NC	NC			
	G2: PhoRTE				NC	NC			
Bick (2021)	G1: VFEs/AB/SOVTs							↑♀: **	
	G2: No therapy								
Desjardins (2022)	G1: VFEs		$\downarrow^{APQ}$ 1.36‡	↓ 1.17‡	$\uparrow^{CPPS}$ 0.68‡	NC			
	G2: VFEs + EMST		NC	NC	$\uparrow^{CPPS}$ 1.03‡	NC			
	G3: VFEs + IMST		NC	↓ 0.87‡	$\uparrow^{CPPS}$ 0.68‡	NC			
Guzman (2018)	G1: WRT					↑***		NC	
	G2: Open vowels (control)					NC		NC	
Inoue (2013)	High-load VFEs	↓1.64	↓1.12	↓0.88					↑1.23
Iwaki (2017)	VFEs	NC	NC	NC					↑**
Kaneko (2015)	G1: VFEs/resonant voice	↓0.58	NC			NC			
	G2: No therapy	NC	NC			NC			
Kaneko (2021)	Vocal resistance training	↓1.21	↓1.10			↑1.39			
Kim (2010)	SKLCT	↓♂:0.80	↓ ♂:0.74 ↓ ♀:1.11	NC				NC	↑ ♂:0.06 ↑ ♀:3.66
Kim (2015)	G1: SKLCT	↓0.94	↓0.71	↓0.56			↑0.24	NC	↑0.78
	G2: VFEs	↓0.48	NC	NC			↑0.42	NC	NC
	G3: No therapy								
Kondo (2022)	WRT/VFEs	↓0.92	NC			NC			
Mato (2018)	VFEs	NC	↓0.60						NC
	Breath-holding + pulling	↓0.56	↓0.32	NC				NC	↑0.40

Study	Interventions	Jitter	Shimmer	NHR	CPP/ Cepstral	SPL	SPL Range	Speaking F_0	F_0 Range
Motohashi (2022)									
Nambu (2019)	VFEs/AB/SOVT/decon-striction	NC	↓0.74	↓0.51					
Sauder (2010) †	VFEs	NC	NC	NC				NC	
Zhu (2018)	Breath-holding + hard onset	↓0.54	↓3.34	↑^{NNE}2.28				↑0.45	

†Alpha < 10. ‡*P* not reported. ***P* < 0.01, ****P* < 0.001.
APQ, amplitude perturbation quotient; CPSS, smoothed cepstral peak prominence in reading; NC, no change.

Appendix B4: Expert-Rated Perceptual Voice Outcomes

Study	Interventions	Overall Severity/Grade	GRBAS Total	Roughness	Breathiness	Asthenia	Strain	Instability	Loudness
Bick (2021)	G1: VFEs/AB/SOVTs		↓***						
	G2: No therapy								
Desjardins (2022)	G1: VFEs	NC							
	G2: VFEs + EMST	NC							
	G3: VFEs + IMST	↓ 0.89‡							
Godoy (2019)	G1: Intensive VTE	↓***		^{SV}↓**	↓***	NC	NC	↓***	^{SV}↑***
	G2: Conventional VTE								
Iwaki (2017)	VFEs	↓*							^{CS}↑*
Kaneko (2015)	G1: VFEs/resonant voice		↓1.64						
	G2: No therapy		NC						
Kim (2015)	G1: SKLCT	↓1.23		↓0.83	↓0.83	NC	NC		
	G2: VFEs	↓1.16		NC	NC	NC	NC		
	G3: No therapy								
Kondo (2022)	WRT/VFEs	↓0.88	↓2.39	↓1.16	↓1.09	↓*	↓0.58		
Mato (2018)	VFEs	↓0.78							
Nambu (2019)	VFEs/AB/SOVT/decon-striction	↓*		↓*	↓*	↓*	NC		
Santos (2014)	G1: WRT	↓*		↓**	NC	↓*	↓**	↓**	
	G2: Vocal education	NC		NC	NC	NC	NC	NC	
Sauder (2010) †	VFEs	NC			^{CS}↓0.67		^{CS}↓ 0.36		
Zhu (2018)	Breath-holding + hard onset	↓3.73^{W}		↓0.89^{W}	↓0.91^{W}	↓0.90^{W}			

†Alpha < 10. ‡*P* not reported.
**P* < 0.05, ***P* < 0.01, ****P* < 0.001.
CS, connected speech; NC, no change; SV, sustained vowel; W, Wilcoxon signed rank effect sixe.

Appendix B5: Glottal Characteristics/Laryngeal Function

Study	Interventions	CQ, Closed Phase	Glottic Gap (NGG, etc)	Bowing	Hyperfunction	Laryngeal EMG	Vibratory Characteristics
Desjardins (2022) ‡	G1: VFEs			↓1.00‡	↓‡		
	G2: VFEs + EMST			↓ 1.01‡	NC		
	G3: VFEs + IMST			NC	↓‡		
Godoy (2019)	G1: Intensive VTE		NC	↓*	NC		NC
			NC	NC	NC		NC

Study	Interventions	CQ, Closed Phase	Glottic Gap (NGG, etc)	Bowing	Hyperfunction	Laryngeal EMG	Vibratory Characteristics
	G2: Conventional VTE						
Gorman (2008)	VFEs		CMF ↓ 0.56				
Guzman (2018)	G1: WRT	↑***					
	G2: Open vowels (control)	NC					
Kaneko (2015)	G1: VFEs/resonant voice		↓ 0.94	NC			a ↑ 1.72
	G2: No therapy		NC	NC			NC
Kaneko (2021)	Vocal resistance training			NC		↑1.69	a ↑1.26
Kim (2010)	SKLCT		↓		↓		
Kim (2015)	G1: SKLCT		↓*		↓*		abc ↑*
	G2: VFEs		↓*		NC		NC
	G3: No therapy						
Mato (2018)	VFEs		↓0.78				
Sauder (2010)	VFEs	NC	NC		NC		

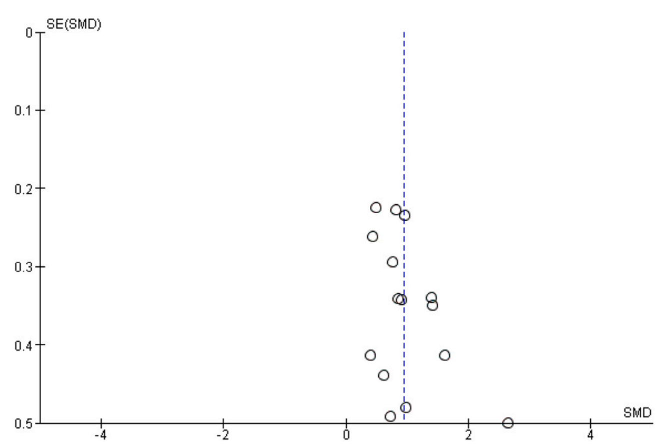
‡*P* not reported. *$P < 0.05$.
CMF, comfortable loudness; NC, no change. Vibratory factors: a) mucosal wave amplitude, b) symmetry, c) regularity.

Appendix B6: Respiratory Outcome Measure Effect Sizes

Study	Interventions	MIP	MEP	Pulmonary Function
Belsky (2021)	G1: PhoRTE + EMST		↑ 3.44	
	G2: PhoRTE		NC	
Desjardins (2022)	G1: VFEs	↑ 0.93‡	↑ 0.47‡	NC
	G2: VFEs + EMST	NC	↑ 0.77‡	NC
	G3: VFEs + IMST	↑ 0.25‡	↑ 0.17‡	NC
Iwaki (2017)	VFEs			NC
Santos (2014)	G1: WRT			↑VC 1.38
	G2: Vocal education			↓VC 0.38

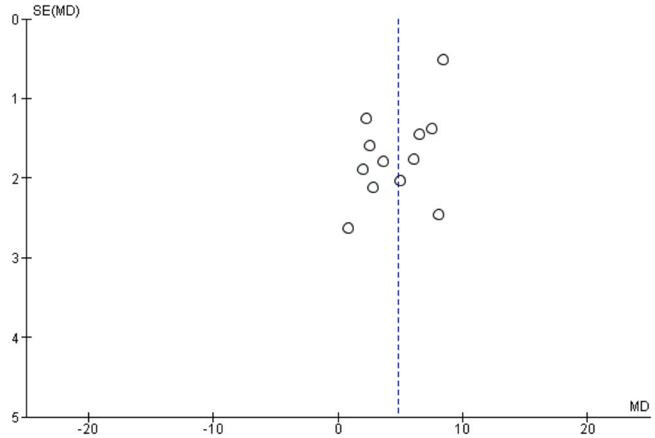
‡*P* not reported.
EMST/IMST, expiratory muscle strength training/inspiratory muscle strength training; NC, no change; PhoRTE, phonation resistance training exercises; VFEs, vocal function exercises; VC, vital capacity; WRT, water resistance therapy.

Appendix C1: Funnel Plot for PROMs Meta-analysis



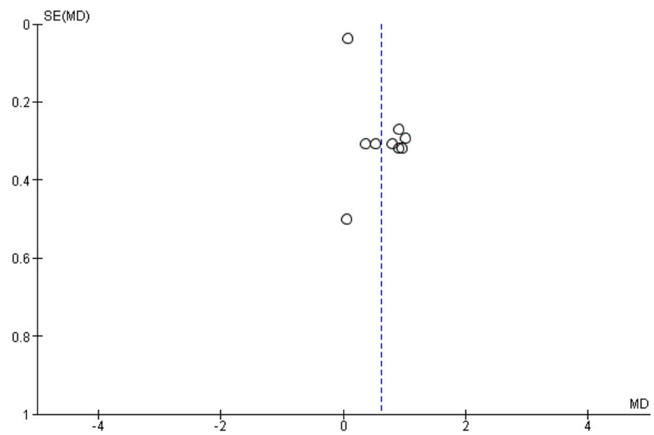
.

Appendix C2: Funnel Plot for MPT Meta-analysis



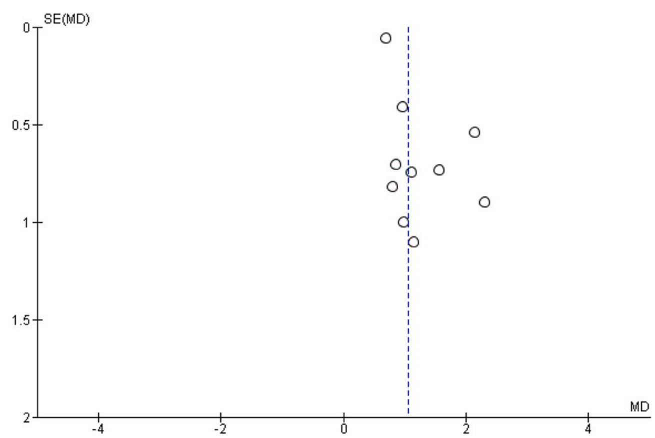
.

Appendix C3: Funnel Plot for Jitter Meta-analysis



.

Appendix C4: Funnel Plot for Shimmer Meta-analysis



.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2023.12.010.

References

1. Golub JS, Chen PH, Otto KJ, et al. Prevalence of perceived dysphonia in a geriatric population. *J Am Geriatr Soc.* 2006;54(11):1736–1739. https://doi.org/10.1111/J.1532-5415.2006.00915.X.
2. Roy N, Stemple J, Merrill RM, et al. Epidemiology of voice disorders in the elderly: preliminary findings. *Laryngoscope.* 2007;117(4):628–633. https://doi.org/10.1097/MLG.0B013E3180306DA1.
3. Wong HY, Ma EP. Self-perceived voice problems in a nontreatment seeking older population in Hong Kong. (Published online). *J Voice.* 2019;35(4):597–603. https://doi.org/10.1016/j.jvoice.2019.12.012. (Published online).
4. Pernambuco L, Espelt A, Góis ACB, et al. Voice disorders in older adults living in nursing homes: prevalence and associated factors. *J Voice.* 2017;31(4):510.e15–510.e21. https://doi.org/10.1016/j.jvoice.2016.11.015.
5. Office for National Statistics. National population projections: 2016-based - Office for National Statistics. Published October 26, 2017. Accessed November 16, 2021. ⟨https://www.ons.gov.uk/releases/nationalpopulationprojections2016basedstatisticalbulletin⟩.
6. Davids T, Klein AM, Johns MM. Current dysphonia trends in patients over the age of 65: Is vocal atrophy becoming more prevalent? *Laryngoscope.* 2012;122(2):332–335. https://doi.org/10.1002/LARY.22397.
7. Jacks A, Kavookjian H, Kraft S. Comparative evaluation and management of dysphonia between adults < 65 and ≥65 years of age. *Otolaryngol Head Neck Surg (United States).* 2021;165(1):142–148. https://doi.org/10.1177/0194599820978435.
8. Takano S, Kimura M, Nito T, et al. Clinical analysis of presbylarynxâ€"Vocal fold atrophy in elderly individuals. *Auris Nasus Larynx.* 2009;37:461–464. https://doi.org/10.1016/j.anl.2009.11.013.
9. Sund LT, Cameron B, Johns MM, et al. Laryngologists' reported decision-making in presbyphonia treatment. (Published online). *J Voice.* 2021. https://doi.org/10.1016/j.jvoice.2021.10.008.(Published online).
10. Johnson AM. Exercising the aging voice. *Perspect ASHA Spec Interest Groups.* 2018;3(3):12–18. https://doi.org/10.1044/persp3.sig3.12.
11. Oates JM. Treatment of dysphonia in older people: the role of the speech therapist. *Curr Opin Otolaryngol Head Neck Surg.* 2014;22(6):477–486. https://doi.org/10.1097/MOO.0000000000000109.
12. Mallick AS, Garas G, McGlashan J. Presbylaryngis: a state-of-the-art review. *Curr Opin Otolaryngol Head Neck Surg.* 2019;27(3):168–177. https://doi.org/10.1097/MOO.0000000000000540.
13. Daniel Lucas PM, Andrea WB. Voice speech therapy in the aged: exploratory systematic review of literature. Atuação Fonoaudiológica na Voz do Idoso: Revisão Sistemática Exploratória de Literatura. Actuación Fonoaudiológica en la Voz de los Ancianos: Revisión Sistemática Exploratoria de la Literatura. *Distúrbios da Comunicação.* 2015;27:2.
14. Galluzzi F, Garavello W. The aging voice: a systematic review of presbyphonia. *Eur Geriatr Med.* 2018;9:559–570. https://doi.org/10.1007/s41999-018-0095-6.
15. Bhatt NK, Garber D, Baertsch H, et al. Treatments for age-related vocal atrophy: a systematic review. Published online March 13 *Laryngoscope.* 2023;133(11):2846–2855. https://doi.org/10.1002/lary.30653.
16. Leão RL de S, Gomes A de OC, Queiroz MRG, et al. Voice therapy with a respiratory approach in older people: an integrative literature review. *Rev CEFAC.* 2022;24(2):e7721. https://doi.org/10.1590/1982-0216/20222427721.
17. Meulenbroek P, Ness B, Lemoncello R, et al. Social communication following traumatic brain injury part 2: Identifying effective treatment ingredients. *Int J Speech Lang Patho.* 2019;21(2):128–142. https://doi.org/10.1080/17549507.2019.1583281.
18. Hart T, Dijkers MP, Whyte J, et al. A theory-driven system for the specification of rehabilitation treatments. *Arch Phys Med Rehabil.* 2019;100(1):172–180. https://doi.org/10.1016/j.apmr.2018.09.109.
19. Hart T., Whyte J., Dijkers M., et al. Manual Of Rehabilitation Treatment Specification. Accessed July 19, 2023. ⟨https://acrm.org/acrm-communities/rehabilitation-treatment-specification/manual-for-rehabilitation-treatment-specification/⟩.
20. Van Stan JH, Whyte J, Duffy JR, et al. Voice therapy according to the rehabilitation treatment specification system: expert consensus ingredients and targets. *Am J Speech Lang Pathol.* 2021;30(5):2169–2201. https://doi.org/10.1044/2021_AJSLP-21-00076.
21. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. Published online March 29 *BMJ.* 2021;372:n71. https://doi.org/10.1136/bmj.n71.
22. Kost KM, Sataloff RT. The aging voice. *Curr Otorhinolaryngol Rep.* 2020;8(3):321–328. https://doi.org/10.1007/s40136-020-00285-8.
23. Kosztyła-Hojna B, Zdrojkowski M, Duchnowska E. Presbyphonia as an Individual Process of Voice Change. *J Voice.* 2023;37(2):303.e1–303.e14. https://doi.org/10.1016/j.jvoice.2020.12.046.
24. Crawley BK, Dehom S, Thiel C, et al. Assessment of clinical and social characteristics that distinguish presbylaryngis from pathologic presbyphonia in elderly individuals. *JAMA Otolaryngol Head Neck Surg.* 2018;144(7):566–571. https://doi.org/10.1001/jamaoto.2018.0409.
25. Lalley PM. The aging respiratory system—Pulmonary structure, function and neural control. *Respir Physiol Neurobiol.* 2013;187(3):199–210. https://doi.org/10.1016/j.resp.2013.03.012.
26. Sterne JAC, Savović J, Page MJ, et al. RoB 2: a revised tool for assessing risk of bias in randomised trials. Published online August 28 *BMJ.* 2019;366:l4898. https://doi.org/10.1136/bmj.l4898.
27. Slim K, Nini E, Forestier D, et al. Methodological Index for Non-randomized Studies (MINORS): development and validation of a new instrument. *ANZ J Surg.* 2003;73(9):712–716.
28. Wan X, Wang W, Liu J, et al. Estimating the sample mean and standard deviation from the sample size, median, range and/or interquartile range. *BMC Med Res Methodol.* 2014;14(1):135. https://doi.org/10.1186/1471-2288-14-135.
29. Higgins J.P.T., Thomas J., Chandler J., et al., eds. *Cochrane Handbook for Systematic Reviews of Interventions*. verion 6.3.; 2022. Accessed August 15, 2023. ⟨https://training.cochrane.org/handbook/current⟩.
30. Taylor C. Range Rule for Standard Deviation. ThoughtCo. Published April 5, 2023. Accessed August 15, 2023. ⟨https://www.thoughtco.com/range-rule-for-standard-deviation-3126231⟩.
31. Johnson AM, Pukin F, Krishna V, et al. Feasibility and preliminary efficacy of two technology-assisted vocal interventions for older adults living in a residential facility. (Published online August). *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.07.019.
32. Moon S, Park J, Yang S. The Effects of Therapeutic Singing on Vocal Functions of the Elderly: A Study on Korean Elderly. *J Voice.* 2022;36(3):437.e1–437.e9. https://doi.org/10.1016/j.jvoice.2020.06.007.
33. Piragibe PC, Silverio KCA, Dassie-Leite AP, et al. Comparison of the immediate effect of voiced oral high-frequency oscillation and flow phonation with resonance tube in vocally-healthy elderly women. Comparação do impacto imediato das técnicas de oscilação oral de alta frequência sonorizada e sopro sonorizado com tubo de ressonância em idosas vocalmente saudáveis. *CODAS.* 2020;32(4):e20190074. https://doi.org/10.1590/2317-1782/20192019074.
34. Segall LE. The effect of group singing on the voice and swallow function of healthy, sedentary, older adults: a pilot study. *Arts Psychother.* 2017;55:40–45. https://doi.org/10.1016/j.aip.2017.02.007.
35. Siracusa M das G de P, Oliveira G, Madazio G, et al. Efeito imediato do exercício de sopro sonorizado na voz do idoso. *J Soc Bras*

Fonoaudiol. 2011;23(1):27–31. https://doi.org/10.1590/S2179-64912011000100008.
36. Tay EYL, Phyland DJ, Oates J. The effect of vocal function exercises on the voices of aging community choral singers. *J Voice.* 2012;26(5):672.e19–672.e27. https://doi.org/10.1016/j.jvoice.2011.12.014.
37. Wu CH, Chan RW. Effects of a 6-week straw phonation in water exercise program on the aging voice. *J Speech Lang Hear Res.* 2020;63(4):1018–1032. https://doi.org/10.1044/2020_JSLHR-19-00124.
38. Liou HH, Hsieh MHC, Hung DSY, et al. The additive effectiveness of inspiratory muscle training on glottic closure and subjective voice outcomes of patients with benign lesion after hyaluronic acid laryngoplasty. (Published online December). *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.10.021.
39. Cho NB, Cho SR, Choi SH, et al. Short-term and long-term efficacy of oropharyngolaryngeal strengthening training on voice using a mobile healthcare application in elderly women [모바일 헬스케어 애플리케이션을 이용한 구강 및 인후두 근력 강화 훈련이 노년 여성의 음성에 미치는 장단기 효능]. *Commun Sci Disord.* 2021;26(1):219–230. https://doi.org/10.12963/csd.21799.
40. Imtiaz R, Ashraf A, Bari S, et al. Presbyphonia: quality of life following voice therapy intervention: presbyphonia through voice therapy intervention. Published online November 30 *Pak J Health Sci.* 2022;3(6):170–174. https://doi.org/10.54393/pjhs.v3i06.351.
41. Mora R, Jankowska B, Guastini L, et al. Computerized voice therapy in hypofunctional dysphonia. *J Otolaryngol.* 2010;39:5.
42. Awan SN, Gartner-Schmidt JL, Timmons LK, et al. Effects of a variably occluded face mask on the aerodynamic and acoustic characteristics of connected speech in patients with and without voice disorders. *J Voice.* 2018;33(5):809.e1–809.e10. https://doi.org/10.1016/j.jvoice.2018.03.002. (Published online).
43. Fabron EMG, Silvério KCA, Berretin-Felix G, et al. Voice therapy for the elderly with progression of intensity, frequency, and phonation time: case reports. Terapia vocal para idosos com progressão de intensidade, frequência e duração do tempo de fonação: estudo de casos. *CoDAS.* 2018;30(6):e20170224. https://doi.org/10.1590/2317-1782/20182017224. Published 2018 Oct 29.
44. Godoy JF, Silverio KCA, Andrade EC de, et al. Intensive voice therapy for the elderly. *Audiol Commun Res.* 2020;25:e2098.
45. Yamaguchi H, Yotsukura Y, Sata H, et al. Pushing exercise program to correct glottal incompetence. *J Voice.* 1993;7(3):250–256. https://doi.org/10.1016/S0892-1997(05)80334-0.
46. Bertelsen C, Zhou S, Hapner ER, et al. Sociodemographic characteristics and treatment response among aging adults with voice disorders in the United States. *JAMA Otolaryngol Head Neck Surg.* 2018;144(8):719. https://doi.org/10.1001/jamaoto.2018.0980.
47. Bradley JP, Hapner E, Johns MM. What is the optimal treatment for presbyphonia? *Laryngoscope.* 2014;124(11):2439–2440. https://doi.org/10.1002/lary.24642.
48. Schuster M, Berghaus A. Therapy methods for senescence effects on speech. *MMW. Fortschritte der Medizin.* 2013;155(1):50–52.
49. Nam IC, Kim SY, Joo YH, et al. Effects of voice therapy using the lip trill technique in patients with glottal gap. (Published online). *J Voice.* 2018;33(6):949.e11–949.e19. https://doi.org/10.1016/j.jvoice.2018.07.013. (Published online).
50. Lombard LE, Steinhauer KM. A novel treatment for hypophonic voice: twang therapy. *J Voice.* 2007;21(3):294–299. https://doi.org/10.1016/j.jvoice.2005.12.006.
51. Dastolfo C, Gartner-Schmidt J, Yu L, et al. Aerodynamic outcomes of four common voice disorders: moving toward disorder-specific assessment. *J Voice.* 2016;30(3):301–307. https://doi.org/10.1016/j.jvoice.2015.03.017.
52. Lin FC, Chien HY, Chen SH, et al. Voice therapy for benign voice disorders in the elderly: A randomized controlled trial comparing telepractice and conventional face-to-face therapy. *J Speech Lang Hear Res.* 2020;63(7):2132–2140. https://doi.org/10.1044/2020_JSLHR-19-00364.
53. Gartner-Schmidt J, Rosen C. Treatment success for age-related vocal fold atrophy. *Laryngoscope.* 2011;121(3):585–589. https://doi.org/10.1002/lary.21122.
54. LaGorio LA, Carnaby-Mann GD, Crary MA. Treatment of vocal fold bowing using neuromuscular electrical stimulation. *Arch Otolaryngol Head Neck Surg.* 2010;136(4):398–403. https://doi.org/10.1001/archoto.2010.33.
55. Gugatschka M, Feiner M, Mayr W, et al. Functional electrical stimulation for presbyphonia: a prospective randomized trial. *Laryngoscope.* 2020;130(11):E662–E666. https://doi.org/10.1002/lary.28489.
56. National Health and Medical Research Council. A Guide to the Development, Implementation and Evaluation of Clinical Practice Guidelines. Published online 1999. Accessed 28 December 2023. https://www.nhmrc.gov.au/sites/default/files/images/a-guide-to-the-development-and-evaluation-of-clinical-practice-guidelines.pdf.
57. Kim ST. Effects of SKLCT for voice improvement in patients with presbyphonia. *Phon Speech Sci.* 2015;7(3):183–191. https://doi.org/10.13064/KSSS.2015.7.3.183.
58. Belsky MA, Shelly S, Rothenberger SD, et al. Phonation resistance training exercises (PhoRTE) with and without expiratory muscle strength training (EMST) for patients with presbyphonia: a non-inferiority randomized clinical trial. *J Voice.* 2021;37(3):398–409. https://doi.org/10.1016/j.jvoice.2021.02.015. (Published online).
59. Berg EE, Hapner E, Klein A, et al. Voice therapy improves quality of life in age-related dysphonia: a case-control study. *J Voice.* 2008;22(1):70–74. https://doi.org/10.1016/j.jvoice.2006.09.002.
60. Bick E, Dumberger LD, Farquhar DR, et al. Does voice therapy improve vocal outcomes in vocal fold atrophy? *Ann Otol Rhinol Laryngol.* 2021;130(6):602–608. https://doi.org/10.1177/0003489420952464.
61. Desjardins M, Halstead L, Simpson A, et al. Respiratory muscle strength training to improve vocal function in patients with presbyphonia. *J Voice.* 2022;36(3):344–360. https://doi.org/10.1016/j.jvoice.2020.06.006.
62. Fujimaki Y, Tsunoda K, Kobayashi R, et al. Independent exercise for glottal incompetence to improve vocal problems and prevent aspiration pneumonia in the elderly: a randomized controlled trial. *Clin Rehabil.* 2017;31(8):1049–1056. https://doi.org/10.1177/0269215516673208.
63. Godoy J, Silverio K, Brasolotto A. Effectiveness of vocal therapy for the elderly when applying conventional and intensive approaches: a randomized clinical trial. *J Voice.* 2019;33(5):809.e19–809.e26. https://doi.org/10.1016/j.jvoice.2018.03.017.
64. Gorman S, Weinrich B, Lee L, et al. Aerodynamic changes as a result of Vocal Function Exercises in elderly men. *Laryngoscope.* 2008;118(10):1900–1903. https://doi.org/10.1097/MLG.0b013e31817f9822.
65. Guzman M, Saldivar P, Pérez R, et al. Aerodynamic, electroglottographic, and acoustic outcomes after tube phonation in water in elderly subjects. *Folia Phoniatr et Logop.* 2018;70(3-4):149–155. https://doi.org/10.1159/000492326.
66. Inoue S, Tokashiki R. Effects of voice therapy for presbyphonia. Accessed July 5 *PT-OT-ST Channel Online J.* 2023;2(A1).Accessed July 5 ⟨http://ptotst-channel.com/journals/page/47⟩.
67. Iwaki S, Wakui E, Takahashi M, et al. Effectiveness of vocal function exercises for presbyphonia. *Jpn J Logop Phoniatr.* 2017;58(2):152–158. https://doi.org/10.5112/jjlp.58.152.
68. Iwaki S, Wakui E, Takahashi M, et al. Problems in long-term maintenance of efficacy of vocal function exercises on presbyphonia. *Jpn J Logop Phoniatr.* 2021;62(2):116–122. https://doi.org/10.5112/JJLP.62.116.
69. Kaneko M, Hirano S, Tateya I, et al. Multidimensional analysis on the effect of vocal function exercises on aged vocal fold atrophy. *J Voice.* 2015;29:638–644. https://doi.org/10.1016/j.jvoice.2014.10.017.
70. Kaneko M, Sugiyama Y, Fuse S, et al. Physiological effects of voice therapy for aged vocal fold atrophy revealed by EMG study. (Published online). *J Voice.* 2021. https://doi.org/10.1016/j.jvoice.2021.09.006. (Published online).

71. Kim ST. The effects of voice therapy in age-related dysphonia. *Phon Speech Sci.* 2010;2(2):117–121.
72. Kondo K, Motomoto Y, Masanobu M. Effects of voice therapy on age-related vocal fold atrophy. *Jpn J Logop Phoniatr.* 2022;63(4):255–261.
73. Mato S, Miyamoto M, Watanabe I, et al. A study on the impact of vocal function exercise (VFE) on vocal fold atrophy. *Jpn J Logop Phoniatr.* 2018;59(4):311–317. https://doi.org/10.5112/jjlp.59.311.
74. Mau T, Jacobson BH, Garrett CG. Factors associated with voice therapy outcomes in the treatment of presbyphonia. *Laryngoscope.* 2010;120(6):1181–1187. https://doi.org/10.1002/lary.20890.
75. Motohashi R, Tokashiki R, Konomi U, et al. Effectiveness of breath-holding pulling exercise in patients with vocal fold atrophy. (Published online August). *J Voice.* 2022. https://doi.org/10.1016/j.jvoice.2022.06.036.
76. Nambu Y, Nimura Y, Mori Y, et al. Treatment for age-related dysphonia: —Voice therapy for patients with vocal fold atrophy. *Jpn J Logop Phoniatr.* 2019;60(1):43–51. https://doi.org/10.5112/jjlp.60.43.
77. Santos SB dos, Rodrigues SR, Gadenz CD, et al. Verifying the effectiveness of using resonance tubes in voice therapy with elderly people. *Audiol Commun Res.* 2014;19:81–87. https://doi.org/10.1590/S2317-64312014000100014.
78. Sauder C, Roy N, Tanner K, et al. Vocal function exercises for presbylaryngis: a multidimensional assessment of treatment outcomes. *Ann Otol Rhinol Laryngol.* 2010;119(7):460–467. https://doi.org/10.1177/000348941011900706.
79. Zhu M, Wang L, Bai X, et al. Hard onset therapy for functional hypoadduction and presbylaryngis: our experience in 22 patients. *Clin Otolaryngol.* 2018;43(4):1122–1125. https://doi.org/10.1111/coa.13078.
80. Ziegler A, Verdolini Abbott K, Johns M, et al. Preliminary data on two voice therapy interventions in the treatment of presbyphonia. *Laryngoscope.* 2014;124(8):1869–1876. https://doi.org/10.1002/lary.24548.
81. Stemple JC, Lee L, D'Amico B, et al. Efficacy of vocal function exercises as a method of improving voice production. *J Voice.* 1994;8(3):271–278. https://doi.org/10.1016/S0892-1997(05)80299-1.
82. Rosen CA, Lee AS, Osborne J, et al. Development and Validation of the Voice Handicap Index-10. *The Laryngoscope.* 2004;114(9):1549–1556. https://doi.org/10.1097/00005537-200409000-00009.
83. Jacobson BH, Johnson A, Grywalski C, et al. The Voice Handicap Index (VHI). *Am J Speech-Lang Pathol.* 1997;6(3):66–70. https://doi.org/10.1044/1058-0360.0603.66.
84. Hogikyan ND, Sethuraman G. Validation of an instrument to measure voice-related quality of life (V-RQOL). *J Voice.* 1999;13(4):557–569. https://doi.org/10.1016/S0892-1997(99)80010-1.
85. Romak JJ, Orbelo DM, Maragos NE, et al. Correlation of the Voice Handicap Index-10 (VHI-10) and Voice-Related Quality of Life (V-RQOL) in Patients With Dysphonia. *J Voice.* 2014;28(2):237–240. https://doi.org/10.1016/j.jvoice.2013.07.015.
86. Johnson AM, Goldfine A. Intrasubject reliability of maximum phonation time. *J Voice.* 2016;30(6):775.e1–775.e4. https://doi.org/10.1016/j.jvoice.2015.11.019.
87. Pearl Solomon N, Garlitz SJ, Milbrath RL. Respiratory and laryngeal contributions to maximum phonation duration. *J Voice.* 2000;14(3):331–340. https://doi.org/10.1016/S0892-1997(00)80079-X.
88. Omori K, Kojima H, Slavit DH, et al. Vocal fold atrophy: quantitative glottic measurement and vocal function. *Ann Otol Rhinol Laryngol.* 1997;106(7):544–551. https://doi.org/10.1177/000348949710600702.
89. Desjardins M, Bonilha S, Carolina S. The impact of respiratory exercises on voice outcomes: a systematic review of the literature. *J Voice.* 2019;34(6):648.e1–648.e39. https://doi.org/10.1016/j.jvoice.2019.01.011. (Published online).
90. Holt-Lunstead J, Smith TB, Layton BJ. Social relationships and mortality risk: a meta-analytic review. (doi:). *PLoS ONE.* 2010;7(7):e1000316. https://doi.org/10.1371/journal.pmed.1000316.
91. Fournier-Foch J, de Mahieu V. Pacebo's in Voice Research: How to Not Affect the Voice?. Faculté de psychologie et des sciences de l'éducation, Université catholique de Louvain, 2020. Prom.: Maryn, Youri; Dedry, Marie. http://hdl.handle.net/2078.1/thesis:24093.
92. Hirano M, Koike Y, Von Leden H. Maximum phonation time and air usage during phonation. *Folia Phoniatr Logop.* 1968;20(4):185–201. https://doi.org/10.1159/000263198.
93. Sawaya Y, Sato M, Ishizaka M, et al. Maximum phonation time is a useful assessment for older adults requiring long-term care/support. *Phys Ther Res.* 2022;25(1):35–40. https://doi.org/10.1298/ptr.E10152.
94. Maslan J, Leng X, Rees C, et al. Maximum phonation time in healthy older adults. *J Voice.* 2011;25(6):709–713. https://doi.org/10.1016/j.jvoice.2010.10.002.
95. Barsties V, Latoszek B, Watts CR, et al. The maximum phonation time as marker for voice treatment efficacy: a network meta-analysis. *Clin Otolaryngol.* 2023;48(2):130–138. https://doi.org/10.1111/coa.14019.
96. Vaca M, Mora E, Cobeta I. The aging voice: influence of respiratory and laryngeal changes. *Orig Res-Laryngol Neurolaryngol Otolaryngol Head Neck Surg.* 2015;153(3):409–413. https://doi.org/10.1177/0194599815592373.
97. Desjardins M, Halstead L, Simpson A, et al. The impact of respiratory function on voice in patients with presbyphonia. (Published online). *J Voice.* 2020;36(2):256–271. https://doi.org/10.1016/j.jvoice.2020.05.027. (Published online).
98. Mezzedimi C, Di Francesco M, Livi W, et al. Objective Evaluation of Presbyphonia: Spectroacoustic Study on 142 Patients with Praat. *J Voice.* 2017;31(2):257.e25–257.e32. https://doi.org/10.1016/j.jvoice.2016.05.022.
99. Johnson AM, Sandage MJ, York N. Exercise science and the vocalist. *J Voice.* 2021;35:376–385. https://doi.org/10.1016/j.jvoice.2019.09.007.
100. Stemple JC. A holistic approach to voice therapy. *Semin Speech Lang.* 2005;26(2):131–137. https://doi.org/10.1055/s-2005-871209.
101. Bode FR, Dosman J, Martin RR, et al. Age and sex differences in lung elasticity, and in closing capacity in nonsmokers. *J Appl Physiol.* 1976;41(2):129–135. https://doi.org/10.1152/jappl.1976.41.2.129.
102. Enright PL, Kronmal RA, Manolio TA, et al. Respiratory muscle strength in the elderly: Correlates and reference values. *Am J Respir Crit Care Med.* 1994;149(2 I):430–438. https://doi.org/10.1164/ajrccm.149.2.8306041.
103. Heman-Ackah YD, Michael DD, Baroody MM, et al. Cepstral peak prominence: a more reliable measure of dysphonia. *Ann Otol Rhinol Laryngol.* 2003;112(4):324–333. https://doi.org/10.1177/000348940311200406.
104. Patel RR, Awan SN, Barkmeier-Kraemer J, et al. Recommended protocols for instrumental assessment of voice: American speech-language-hearing association expert panel to develop a protocol for instrumental assessment of vocal function. (Published online). *Am J Speech-Lang Pathol.* 2018;27(3):887–905. https://doi.org/10.1044/2018_AJSLP-17-0009. (Published online).