Tracheoesophageal Voice Therapy in Postlaryngectomy Rehabilitation: A Systematic Review

*,†Freya Sparks, ‡Margaret Coffey, *Lucy Dipper, *Sally Morgan, and *Katerina Hilari, *†‡*London, United Kingdom*

Summary: Background. Following total laryngectomy, surgical voice restoration is considered the optimal modality for re-establishing communication via tracheoesophageal voice. Yet beyond the insertion of a voice prosthesis to elicit voice production, there is suboptimal clinical knowledge of how to rehabilitate the perceptual quality of tracheoesophageal voice. This systematic review will identify and critically evaluate the quality and effectiveness of therapeutic interventions for tracheoesophageal voice. The findings of this review will inform the development of a novel tracheoesophageal voice therapy intervention.

Study design. Systematic literature review carried out in accordance with PRISMA guidelines.

Methods. The review protocol was registered with PROSPERO. Eight electronic databases were searched using a prespecified search strategy. Records were independently screened by two reviewers against inclusion and exclusion criteria. Eligible studies were assessed for quality using the PEDro, ROBIN-T, and NHLBI critical appraisal tools. Data was extracted pertaining to participant characteristics and the content, dosage, intensity and outcomes of interventions.

Results. 6344 records were identified, of which 38 were included for full-text review. Six studies met the eligibility criteria for inclusion. Voice rehabilitation was not the primary focus in the majority of studies, and the risk of bias was identified across studies. There was significant heterogeneity in the interventions and outcome measures used within studies with insufficient detail provided on intervention content for tracheoesophageal voice. Evidence for the effectiveness of interventions was limited and inconsistent across studies.

Conclusions. This review found that tracheoesophageal voice therapy is an under-researched area of clinical practice. Evidence from the small body of existing studies was not sufficiently robust to inform clinical practice at this time. This review highlights the necessity to develop and test interventions aimed at improving the perceptual quality of tracheoesophageal voice.

Key Words: Tracheoesophageal–Laryngectomy–Voice–Alaryngeal–Rehabilitation.

INTRODUCTION AND OBJECTIVES

Total laryngectomy has devastating and irreversible consequences for communication, resulting in the permanent loss of the ability to use a conventional voice to speak and communicate. Surgical voice restoration is commonly offered as the optimal treatment for re-establishing spoken communication with tracheoesophageal voice, via the insertion of a voice prosthesis.[1] However, there is currently no professional consensus or systematic review that addresses the specific question of how to improve tracheoesophageal voice quality after total laryngectomy. Clinical knowledge is therefore limited on what therapy approaches should be used to improve tracheoesophageal voice quality and maximise functional communication.

Laryngectomy is normally offered with curative intent[2,3] with over 30% of the people with advanced laryngeal cancer surviving for five years or more.[4] As a greater number of people are cured of cancer and enter survivorship, the need for lifelong rehabilitation is essential.[5]

National guidelines for head and neck cancer rehabilitation[6] stipulate that long-term speech and language therapy (SLT) is essential for people undergoing total laryngectomy, and surgical voice restoration should be offered to all people with laryngectomy. To support this provision, it is the recommendation of the British Association of Head and Neck Oncologists (BAHNO) that an SLT with specialist surgical voice restoration skills is present in all head and neck cancer units.[7] Whilst existing guidelines highlight the importance of the SLT role in laryngectomy rehabilitation, detail on the recommended content of SLT intervention is insufficient. In the context of tracheoesophageal voice rehabilitation, professional guidance from the Royal College of Speech and Language Therapists (RCSLT),[8] centres on the insertion and maintenance of a voice prosthesis. There is not recognition that tracheoesophageal voice quality may need to be optimised and problems may extend beyond the initial difficulty establishing voice.

Sharpe et al (2019)[9] identified that patients reported quality of life was affected by changes to communication after total laryngectomy and suggested that linking

Accepted for publication October 27, 2023.

Presentation: Interim data from this systematic review was presented at the British Association of Head and Neck Oncologists annual conference: Sparks, F. Coffey, M. Dipper, L. Morgan, S.Hilari, K. (2022). A systematic review of tracheoesophageal voice therapy inpost laryngectomy rehabilitation. Oral presentations. British Association of Head and Neck Oncologists (BAHNO) conference. 18th May, London, UK.

From the *Department of Language and Communication Sciences, City, University of London, London, United Kingdom; †Speech and Language Therapy Department. Barts Health NHS Trust, London, United Kingdom; and the ‡Speech and Language Therapy Department. Imperial College Academic Health Science Centre, London, United Kingdom.

Address correspondence and reprint requests to Freya Sparks, Department of Language and Communication Sciences, School of Health and Psychological Sciences, City, University of London, Northampton Square, London EC1V 0HB. E-mail: Freya.sparks@city.ac.uk

Journal of Voice, Vol xx, No xx, pp. xxx–xxx

0892-1997

https://doi.org/10.1016/j.jvoice.2023.10.033

communication to quality of life is a useful approach in post laryngectomy management. Mental health disorders such as depression and anxiety[10,11] affect around 22% of people after laryngectomy[12] and their presence has been linked to the outcome of voice restoration.[12] People with head and neck cancer have a three-fold increase in incidence of suicide compared to the general population in North America and within head and neck cancer, laryngeal cancer is among the highest incidence.[13] It could therefore be interpreted that intervention aimed at improving communication is a salient contributor to improving quality of life and supporting mental health after laryngectomy.

Several systematic reviews have been completed within the area of postlaryngectomy voice rehabilitation. These reviews have not focussed on *therapeutic* interventions, however, and instead evaluate the outcomes of different surgical approaches or comparison of postlaryngectomy communication methods.[14–18]

The psychosocial factors which influence voice rehabilitation outcomes were evaluated in a systematic review by Singer et al (2007).[19] This review sought to identify the factors associated with the success of voice rehabilitation after laryngectomy, noting that *'there is no overall accepted criterion for success in gaining a new voice after laryngectomy'*.[19] Employment status, physical condition, communication method and behaviour were associated with rehabilitation outcomes. However, the authors noted that study results were inconsistent and the lack of standardised, validated measures and criteria for success presented a challenge in comparing studies and drawing conclusions. These areas were therefore recommended for further study.

The effect of different prostheses on voice quality has been evaluated.[20] This study did not find a specific optimal prosthesis based on expert rating; however, individual patient preference was identified as a salient factor and thus patient involvement in decision-making was advised. Tawfik et al (2021)[21] conducted a large systematic review and meta-analysis of the outcome of voice prothesis usage. The aim of this review was to compare different voice prostheses and to identify which type of prosthesis was superior across various parameters. The review therefore centred on voice restoration, rather than therapeutic interventions. The review concluded that the Provox-2 was the best voice prosthesis in terms of airflow, sizing, patient preference, and complications.

In summary, the focus of the literature on surgical voice restoration has been towards the placement of protheses, prosthesis selection, and management of complications, rather than on improving the perceptual quality of tracheoesophageal voice. There is, therefore, a gap in knowledge around the evidence for therapeutic interventions that optimise tracheoesophageal voice quality.

The objective of this systematic review was to critically appraise and synthesise existing evidence on *therapeutic* interventions that target improvement in tracheoesophageal voice quality. The review will inform clinical practice and the development of a new therapy approach for tracheoesophageal voice users. Specifically, the review aims to:

1. Identify and describe the evidence pertaining to therapeutic interventions for tracheoesophageal voice after total laryngectomy
2. Critically appraise any existing therapeutic interventions, evaluate their effectiveness and the quality of the reported evidence

METHODS

The review was conducted in accordance with the preferred reporting items for systematic reviews and meta-analyses (PRISMA) guideline.[22] Further information is presented in the PRISMA checklist (Supplementary Material A). The protocol was registered with the International Prospective Register of Systematic Reviews (PROSPERO) under registration reference: CRD42021265095. PROSPERO and Cochrane databases were searched to ensure there was no existing review of the same research questions.

Eligibility

The population of interest, and therefore, primary inclusion criteria, was adults with total laryngectomy and tracheoesophageal voice prosthesis. Studies were excluded if they related to sub-total laryngectomy procedures such as partial laryngectomy or tracheostomy. Studies relating to outdated surgical procedures practiced during the evolution of tracheoesophageal fistula surgery were excluded. Studies that recruited mixed populations were included if the findings for the total laryngectomy with tracheoesophageal prosthesis group were reported separately, or if they formed more than 25% of the sample.

Therapeutic intervention to optimise functional voice after total laryngectomy is an emerging area with a limited evidence base. To identify as much evidence as possible, a broad criterion was applied to inclusion. Published, peer-reviewed intervention studies of any design with a comparator or control group were included. Interventions delivered in any setting were included. The search was not restricted by language or date of publication.

Interventions were required to target therapeutic rehabilitation of tracheoesophageal voice after total laryngectomy. To clarify the definition of therapeutic intervention, the Van Stan Taxonomy of Voice Therapy was used.[23] Studies were included if the intervention could be categorised according to the direct and indirect interventions listed within the taxonomy. For example, direct interventions comprised working on respiratory function, vocal function, musculoskeletal aspects such as postural alignment, or auditory and somatosensory features. Indirect interventions comprised of enhancing knowledge and addressing psychosocial factors.

As this is an under-researched area with scarce validated measures, outcomes across different areas were included.

TABLE 1.
Search Strategy for Systematic Review

Search	Terms
1	Laryngectom* OR tracheoesophag* OR alaryngeal OR 'surgical voice restoration'
2	Subject headings: EBSCO: Laryngectomy OR Tracheoesophageal voice prosthesis OR Speech, Alaryngeal OVID: Laryngectomy OR Larynx prosthesis OR Voice prosthesis OR Alaryngeal speech
3	S1 OR S2
4	voice OR communicat* OR speech OR phonat*
5	Subject headings: EBSCO: Voice OR Voice Quality OR Voice disorders OR Communication OR Speech OR Phonation OVID: Voice OR Voice disorder OR Speech OR Phonation
6	S4 OR S5
7	Therap* OR intervention OR treat* OR strateg* OR rehabilitat* OR training OR programme*
8	Subject headings: EBSCO: Rehabilitation OR Rehabilitation, Speech &Language OR Voice Therapy OVID: Therapy OR Rehabilitation OR Speech & Language rehabilitation OR Voice Training
9	S7 OR S8
10	S6 AND S9
11	S3 AND S10

Studies were eligible if their reported outcomes related to voice, speech, communication and their impact across ICF domains, quality of life or well-being. Non-therapeutic interventions were excluded, such as surgery or instrumental interventions as these would not inform what Speech and Language Therapists, specifically, can do in the scope of their practice.

Information sources

Eight databases were searched via Ovid and EbscoHost platforms in July 2021 and updated in February 2023: CINAHL, Medline, Health Policy Reference Centre, APA PsychInfo, Embase, AMED, Cochrane, Ovid Emcare.

Search strategy

The search strategy (Table 1) was developed in consultation with a specialist subject librarian. Databases were searched separately using keywords and subject terms for title and abstract search.

Selection process

Search results were exported into EndNote for deduplication, following which a manual deduplication was completed to identify any missed duplicates. Deduplicated results were uploaded to Covidence software to support the management of the process. An additional deduplication was run in Covidence. The PRISMA flowchart was used as a framework to document each stage of the review (Figure 1).

Due to resource limitations, title and abstract screening was completed in two stages. First, one reviewer (FS) checked the title and abstracts against the primary inclusion criterion of the adult population with total laryngectomy with voice prosthesis. 10% of the papers were checked by a second reviewer (SM). The two reviewers (FS and SM) then independently screened titles and abstracts of the included papers against the full inclusion and exclusion criteria to identify studies for full-text review. Disagreements were resolved by consensus between the reviewers, with a third senior reviewer (KH) available for a final decision if consensus could not be reached. At this stage 12 studies were discussed by the reviewers (FS and SM) to reach a consensus on inclusion for full-text review, senior review was not required. Reference lists and citations of included papers were screened for any studies that may have been missed in the initial search. Full-text review was completed independently by FS and SM using the same process to resolve any disagreements. Again senior review was not required as one inconsistency was resolved between the reviewers (FS and SM).

Data collection process

An electronic data extraction form was created to capture relevant data items from included studies. Data extraction was completed by the first author (FS). A second reviewer (SM) carried out independent data extraction on a randomly selected 30% sample of included papers to ensure consistency. No inconsistencies were identified in data extraction.

Risk of bias assessment

The quality of the included studies was rated independently by two of the review teams (FS and SM). Assessment of quality was undertaken at the study level. The following quality assessment tools were used according to the study design:

- For before and after studies with no control group, the National Heart, Lung and Blood Institute of the National Institute for Health quality assessment tool

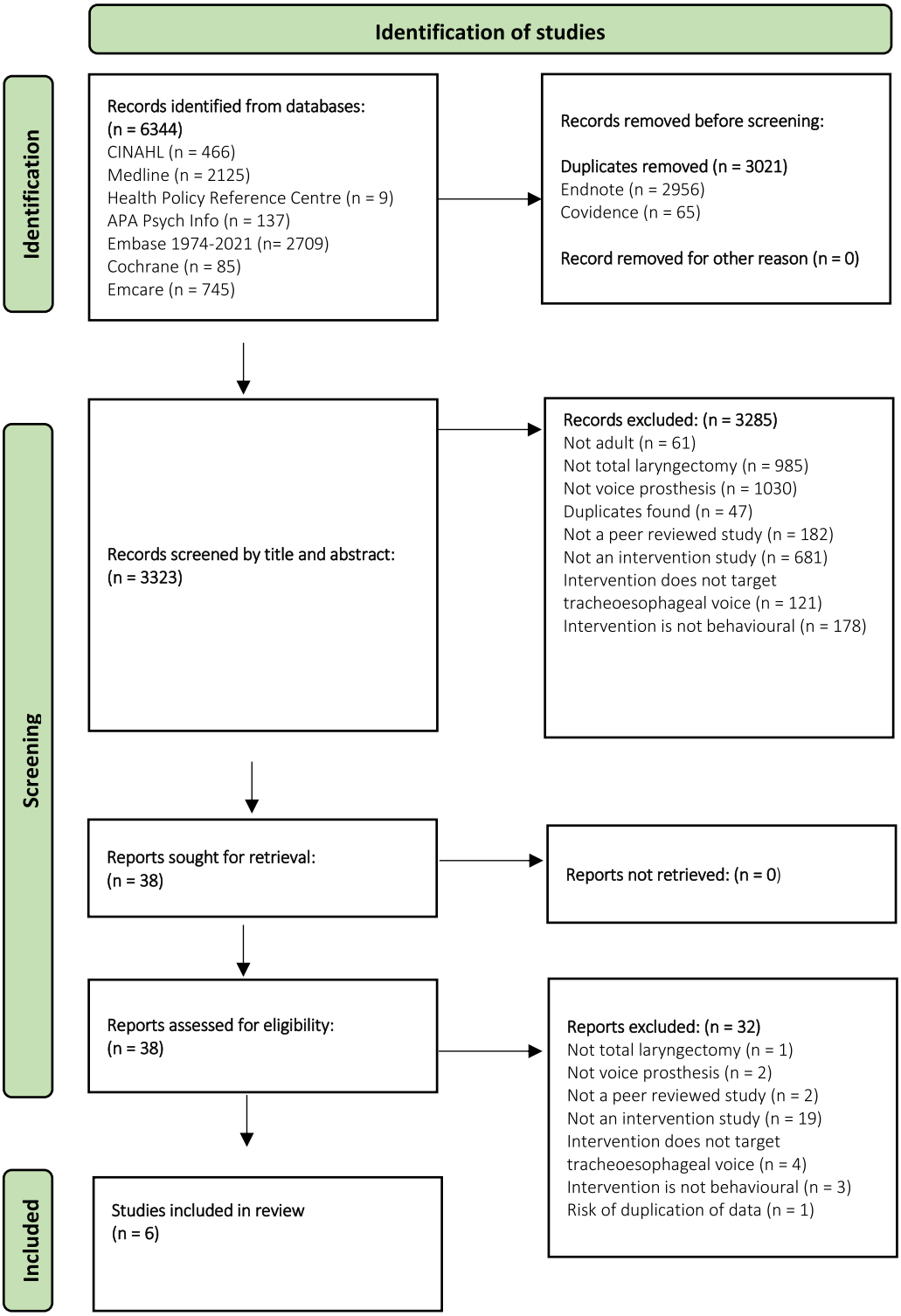

FIGURE 1. Identification of studies. PRISMA 2020 Flow diagram adapted from Page et al, 2021.

for Before-After (Prepost) Studies with No Control Group[24]
- For randomised and non-randomised control trials, the Physiotherapy Evidence Database (PEDro) scale[25]
- For single case studies, the Risk of Bias in N-of-1 Trials (ROBiN-T) scale[26]

The template for intervention description and replication (TIDieR) checklist[27] was used to assess the overall quality of the intervention reporting.

Effect measures

Effect sizes on primary outcome measures were calculated where data were appropriate using Cohen's d.[28] Where studies did not specify primary outcome measures, effect sizes were calculated for the overall score of measures used. Cohen's d effect sizes were calculated using Psychometrica online calculator (Lenhard and Lenhard, 2016). Effect sizes were interpreted using the following values: small (d = 0.2), medium (d = 0.5), large (d = 0.8) according to values advised by Cohen.[29]

A heterogeneity assessment was planned using the I^2 measure to ascertain if it was possible to perform a meta-analysis on relevant group studies; however, this was not performed as no two studies explored the same or very similar interventions. Single case studies were reported separately.

Synthesis methods

Studies of adequate quality, as defined by the quality assessment tools listed above, were included for data synthesis. Descriptive statistics were used to report participant characteristics and other variables. Due to the heterogeneity in the interventions evaluated and outcome measures used, a narrative synthesis approach was implemented, informed by ESRC guidance.[30]

RESULTS

Study selection

Database searching yielded 6344 records. Following deduplication, title and abstract screening, and full-text review, six studies were included for data extraction. Thirty-two studies were excluded following full text review. Citations and reasons for exclusion are listed in Supplementary Material B. Citation checking of included studies found no additional studies that met the inclusion criteria. Figure 1 illustrates the process of study identification. Supplementary Material C details the data extracted from included studies, with the exception of key results and effect sizes which are reported in the results section.

Translation of non-English papers

Title and abstracts (n = 3) and full texts (n = 6) requiring translation were first informally translated to English by author connections or Google Translate. Two studies were deemed to potentially meet criteria for inclusion and therefore formal translation was sought using a medical specialist translation agency.

Study characteristics

Two studies were two-arm randomised controlled trials with intervention versus control groups.[31,32] Three studies used a prospective before and after design, without a control group, where timepoint was used as the comparator.[33–35] These studies did not use multiple baseline data collection points, and only one study[35] used two postintervention data collection points. One study[36] was a single case report with pre and postintervention outcome measures. Three studies took place in The Netherlands,[32,35,36] two in Brazil,[33,34] and one in Italy.[31] Two studies reported that participants carried out the intervention at home.[32,35] Other studies did not give precise detail on the setting of the intervention but were indicative of an outpatient setting.

Participant characteristics

All studies centred on participants with total laryngectomy and either exclusively, or the majority, of participants were voice prosthesis users. Information was provided on extended surgeries or reconstruction and all studies, with the exception of De Oliveira et al, (2005)[34] reported neoadjuvant or adjuvant oncological treatments. The studies reported on 157 participants, of whom 96 had an intervention and 61 were in control groups. Across studies the age of participants ranged from 40 to 81 years (where reported) and the mean age ranged from 49.8 to 68 years (where reported). All studies included a higher percentage of male participants, except the female single case report. All studies targeted alaryngeal voice produced by a prosthesis, however, the single case study participant used tracheojejunal voice in place of tracheoesophageal voice.[36]

Intervention characteristics

Intervention description

Studies followed a set protocol of intervention, however in most, this was not provided and interventions were not described in sufficient detail for replication. Modifications were not covered in any studies and fidelity was covered in two. Reports lacked information on the specifics of the intervention, such as the exact nature of an exercise, the number of repetitions, or the rationale for the exercise. More detail was provided in a separate protocol paper[37] for Jansen (2020)[32] which described the intervention. Other studies used photographs to enhance descriptions.[35,36] The TIDieR checklist[27] describes intervention characteristics for all included studies in Supplementary Material D.

Intervention type

All interventions could be categorised within the Van Stan Taxonomy of Voice Therapy[23] which comprises direct interventions such as musculoskeletal and

TABLE 2.
Risk of Bias by Study Design

Randomised controlled trials: PEDro scale criteria	Longobardi, 2019	Jansen, 2020
1. Eligibility criteria were specified	No	Yes
2. Subjects randomly allocated to groups	Yes	Yes
3. Allocation was concealed	No	Yes
4. Groups similar at baseline in most important prognostic indicators	No	Yes
5. Blinding of all subjects	No	No
6. Blinding of therapists who administered therapy	No	No
7. Blinding of all assessors who measured at least one key outcome	Yes	No
8. Measures of at least one key outcome obtained from more than 85% of subjects initially allocated to groups	Yes	Yes
9. All subjects received treatment or control as allocated, or if not, data for at least one key outcome was analysed by 'intention to treat'	Yes	Yes
10. Results of between-group statistical comparisons are reported for at least one key outcome	Yes	Yes
11. Study provides point measures and measures of variability for at least one key outcome	Yes	Yes
Score	6/10 Good	7/10 Good

continued on next page

respiratory; indirect interventions such as pedagogy and counselling; and intervention delivery methods. The two RCTs used indirect interventions in the form of self-care education programmes focussed on prosthesis and stoma care; and psychological support (Longobardi et al, 2019 only).[31] Within direct interventions, respiratory, musculoskeletal and vocal function interventions were used most frequently. However, approaches differed and a lack of detailed description prevented identification of specific similarities in therapeutic activities. Respiratory interventions were used in two studies[31,34] in the form of breathing-to-stoma occlusion coordination. Whereas Van Sluis et al (2020)[35] and Onofre (2013)[33] used different approaches to respiratory muscle strength training. Musculoskeletal interventions were described in one study[36] using digital modification and postural optimisation. Differing interventions targeting vocal function were used across studies with the exception of Baijens et al (2010)[36] and Van Sluis et al (2020)[35] which did not use this approach. Interventions focussed on articulation, prosody, and pitch modulation, whilst one study[33] used singing activities with the aim of improving auditory-perceptual voice quality. The dosage varied across studies from daily to weekly intervention sessions, and the overall intervention period was heterogenous, ranging from 4 weeks to 7 months.

Intervention delivery

Interventions were delivered via face-to-face sessions. In two studies[32,35] home practice followed an initial face-to-face session to introduce the intervention. All studies used individual sessions with the exception of Longobardi et al (2019)[31] where group sessions were used after an initial block of individual sessions.

Fidelity

Two studies reported adherence measures.[32,35] In both, a practice tracking diary was used alongside informal qualitative measures such as a questionnaire or weekly check-up conversation. Jansen et al (2020)[32] reported percentages across high, medium, and low adherence rather than the more commonly reported percent adherence.

Attrition

No attrition was reported in three studies.[31,34,36] One participant withdrew from Van Sluis et al (2020)[35] due to unrelated medical reasons and one participant died of unrelated causes during the intervention period in Jansen et al, 2020.[32] Onofre (2013)[33] reported that two participants were excluded due to a lack of adherence to the intervention.

TABLE 2. Continued

Before/After studies with no control group: NHLBI scale (yes, no, cannot determine, not reported, not applicable)	Onofre, 2013	De Oliveira, 2005	Van Sluis, 2020
1. Was the study question or objective clearly stated?	Yes	Yes	Yes
2. Were eligibility criteria for the study population prespecified and clearly described?	No	No	Yes
3. Were participants representative of those who would be eligible for the intervention in the clinical population of interest?	Yes	Yes *Predominantly male	No
4. Were eligible participants that met the prespecified entry criteria enrolled?	Cannot determine	Cannot determine	Cannot determine
5. Was the sample size sufficiently large to provide confidence in the findings?	No	No	Yes
6. Was the intervention clearly described and delivered consistently across the study population?	No	No	Yes
7. Were outcome measures prespecified, clearly defined, valid, reliable and assessed consistently across all participants?	No	No	Yes
8. Were the people assessing the outcomes blinded to the participants' interventions?	Not reported	Yes	Not reported
9. Was the loss to follow-up after baseline 20% or less. Were those lost to follow-up accounted for in the analysis?	No	Yes	Yes
10. Did statistical methods examine changes in outcome measures before/after intervention? Were statistical tests done that provided p values for pre/post change?	No	Yes	No
11. Were outcome measures of interest taken multiple times before and after intervention	No	No	No
12. If intervention was conducted at group level did statistical analysis take into account use of individual level data to determine effects at group level?	Not applicable	Not applicable	Not applicable
Overall quality rating	POOR	FAIR	FAIR

continued on next page

Adverse effects

Adverse effects were reported by Van Sluis et al (2020)[35] only. There was one event of voice prosthesis dislodgement after a training session. Additionally, minor complaints of dizziness and poor stoma seal around the EMST device were reported. Participants did not like the instruction to plug their voice prosthesis during EMST practice, however, this was not deemed an essential intervention component.

Outcome measures

No studies used the same outcome measures. Three studies assessed aspects of perceptual voice quality as an outcome, however different outcome measures were used. Several studies used voice quality outcome measures borrowed from laryngeal voice therapy, which were not validated on the laryngectomy population, such as the GRBAS scale,[38] Voice Handicap Index,[39] and Speech

TABLE 2. Continued

Single case study: ROBiN-T scale	Baijens, 2010
1. Design	0
2. Randomisation	0
3. Sampling behaviour during all phases	0
4. Blinding of participant and practitioner delivering the intervention	0
5. Blinding of the assessor	2
6. Inter-rater agreement	0
7. Treatment adherence	0
Internal validity sub-score	**2 / 7**
8. Baseline characteristics of participants	1
9. Setting	0
10. Dependent variable – target behaviour	1
11. Independent variable – intervention	2
12. Raw data record	0
13. Data analysis	1
14. Replication	0
15. Generalisation	0
External validity sub-score	**5 / 16**
Total score	**7 / 30**

Handicap Index.[40] Supplementary Material E details outcome measures used.

Risk of bias

A risk of bias assessment was carried out on all studies. Table 2 reports this by study design.

Risk of bias across studies

Risk of bias was moderate to high with limitations identified in all studies. Across studies sample sizes were relatively small with a higher overall proportion of male participants. The mean average age was towards the lower range of the clinical population in the two studies.[33,34] Few studies incorporated participants with extended surgery. No studies reported ethnicity or socioeconomic status of participants.

Selection bias was identified in several studies due to a lack of stated eligibility criteria, unclear recruitment protocols and absence of concealed allocation in one randomised controlled trial. Blinding of participants and/or researchers was not consistently applied across studies leading to the risk of performance bias. The potential for attrition bias was low, with the exception of Onofre et al, (2013)[33] which presented high risk of attrition bias due to the exclusion of participants with poor adherence to the intervention.

Intervention description was inconsistent in several studies, impacting on replicability, and one study[32] presented a risk of confounding bias due to the intervention design. Outcome measures were heterogenous with reliance on perceptual ratings, informal measures or measures that were not validated within the laryngectomy population, presenting a further risk of performance bias.

Results of individual studies

Table 3 details the results and effect sizes of all studies included for data extraction.

Data synthesis

Due to the heterogeneity of the studies, interventions tested and outcome measures used, meta-analysis was not carried out. A narrative synthesis approach was applied to describe the interventions and the outcomes. Two studies[33,36] were excluded from the data synthesis as they were of insufficient quality for inclusion. The four included studies[31,32,34,35] totalled 151 participants, of which 132 were male and 19 female.

Intervention description and theory

Studies were heterogenous in their interventions, stated aims, and use of outcome measures. There were however common elements targeting voice production between

TABLE 3.
Summary of Results and Effect Sizes of Individual Studies

Study	Longobardi et al, 2019	Baijens et al, 2010	Van Sluis, 2020	De Oliveira, 2005	Jansen, 2020	Onofre, 2013
Results	Two-way ANOVA to compare group with SLT and Psychology intervention to group with SLT intervention only 1. SCL-90: Significant difference between groups ($P < 0.05$) on 5/10 subsections: O-C, DEP and PARA. Significant difference between time for 2/10 subsections: ANX ($P < 0.038$) and HOS ($P = 0.021$) Interaction effects not reported 2. WHO-QOL: Significant difference between groups ($P = 0.017$) in REL domain; and time ($P < 0.004$) in PHY domain. Interaction effects not reported 3. I-SECEL: Significant difference between groups ($P = 0.04$) and time ($P < 0.001$) on total scores. Interaction effects not reported 4. INFVO All subscales showed a significant difference between groups (all *P* values below 0.005) and time (all *P* values below	No statistical tests performed 1a. VHI: Pre-treatment score: 80/120; post treatment score: 59/120 1b. Videofluoroscopy: Pre-treatment: hypotonicity, no constriction of vibratory segment, severe dilatation. Post treatment: increased tonicity and constriction of vibratory segment, more so with digital pressure 1c. High speed endoscopy: Post treatment comparison of vibration with and without digital pressure showed improved mucosal wave during application of digital pressure	No statistical tests reported 1. Pulmonary function Increase in mean MEP score at week 4 vs baseline: (mean/SD): Baseline: 125.5/ (19.6) Week 4: 174.8/ (28.6) No change in PEF, VC, FEV1 or COPD questionnaire 2. CPET No difference in physical exertion outcomes for CPET and Borg scales 3. Short Fatigue Questionnaire No change over time in fatigue 4. VHI-10 and Vocal Function No change over time in VHI-10 score, MPT or Hz. Db increased at week 4 but returned to baseline at week 8 follow up: Mean / SD: Baseline: 26.4 dB / (5.6) Week 4: 31.9 / (3.7) Week 8: 27.6 / (5.4) 5. Feasibility and compliance One participant	Binomial tests performed based on those who improved versus those who did not improve Significant difference reported in the number of speech and voice parameters (23 in total) that improved ($n = 18$) compared to those that declined ($n = 5$) ($P = 0.012$) 11/17 participants improved in at least one area but there was no significant difference in means 1a. Sentence melody No significant difference in mean 1b. Sentence intelligibility No significant difference in mean 1c. Spontaneous speech intelligibility No significant difference in mean 1d. Singing voice pitch and vocal quality No significant difference in mean	Two-way mixed ANOVA. 1. SHI No significant group or time or interaction effects 2a. EORTC QLQ-C30 No significant group, time or interaction effects 2b. EORTC QLQ-H&N35 No significant group, time or interaction effects 3a. Adherence in experimental group High in 49% Moderate in 10% Low in 41% 3b. PAM No significant group, time or interaction effects 4. SWAL-QOL The experimental group performed better with significant interaction effect for overall score ($P = 0.013$) and communication sub-score ($P = 0.004$)	No statistical tests performed 1a. GRBAS Improvement in general degree of dysphonia, roughness and breathiness. Worsening in strain on vowel /i/ 1b. Likert scale of pitch Pitch shifted in three participants postintervention 2a. Semi-tone range Extended semi-tone range reported postintervention, however baseline data is not provided 2b. Likert scale of musicality Presence of tuning and legato increased postintervention

TABLE 3. *(Continued)*

Study	Longobardi et al, 2019	Baijens et al, 2010	Van Sluis, 2020	De Oliveira, 2005	Jansen, 2020	Onofre, 2013
	0.022) for all participants. Interaction effects not reported 5. Intelligibility No significant difference between groups. Significant difference in mean percentage of correctly perceived words over time ($P < 0.001$)		withdrew for unrelated reasons 95.5% compliance with training programme was achieved. Participants reported staying motivated during intervention. 4/9 participants reported air leakage under baseplate, 2/9 participants experienced dizziness during and after training. Participants were unable to plug their voice prosthesis or found this problematic One event of voice prosthesis dislodgement during training			

TABLE 3. *(Continued)*

Study	Longobardi et al, 2019	Baijens et al, 2010	Van Sluis, 2020	De Oliveira, 2005	Jansen, 2020	Onofre, 2013
Effect size of primary outcome measure or all measures if primary outcome not specified, Cohen's d (95% confidence intervals)	Primary outcome measure not specified SCL-90: Experimental group subsection scores pre/ postintervention: O-C: d: −1.14 (−2.16 to −0.11) DEP: d: −1.22 (−2.26 to −0.19) PARA: d: −0.88 (−1.87 to 0.12) WHO-QOL: Social relationship domain score postintervention between groups: d: −1.14 (−1.88 to −0.39) I-SECEL: Overall score postintervention between groups: d: 1.61 (0.81 – 2.41) INFVo Domain scores postintervention between groups: I domain: d: −5.80 (−7.38 to −4.22) N domain: d: −5.08 (−6.5 to −3.65) F domain: d: −1.33 (−2.09 to −0.560 Vo domain: d: −4.49 (−5.79 to −3.19) Intelligibility: Mean/SD not given, unable to calculate	No statistical tests reported	Primary outcome measure not specified Effect sizes calculated for MEP and dB MEP baseline/week 4: d: 2.01 (0.41–3.61) MEP baseline/week 8: d: 1.56 (0.06–3.05) dB range baseline/week 4: d: 1.16 (−0.25 to 2.57) dB range baseline/week 8: d: 0.22 (−1.09 to 1.53)	No statistical tests reported	SWAL-QOL: Overall score postintervention between groups: d: −0.18 (−0.62 to 0.26) Overall score at six months between groups: d: −0.53 (−0.97 to −0.07)	No statistical tests reported

TABLE 3. *(Continued)*

Study	Longobardi et al, 2019	Baijens et al, 2010	Van Sluis, 2020	De Oliveira, 2005	Jansen, 2020	Onofre, 2013
Outcome summary	1. SCL-90 Experimental group showed significant improvement in SCL-90 domains over time, whereas Control group scores worsened in QOL and depression domains. 2. WHO-QOL Mean score in 'social relationships' domain improved over time in the experimental group, whereas the control group score worsened Both groups improved mean scores over time on I-SECEL, INFVO and Intelligibility	Improved score on VHI indicated reduction in self-reported voice-related handicap. Subjective rating of Videofluoroscopy and endoscopy showed improvement in tonicity, constriction and mucosal wave	Increased maximum expiratory pressure, and dynamic range in dB was found after four weeks, however increase in dB was not sustained. Compliance was high and EMST was feasible and safe after laryngectomy. No changes over time to measures of physical exertion, fatigue or other measures of vocal function	Based on composite score there was a significant difference between the number of parameters that improved compared to those that declined ($P = 0.012$), however no significant differences were found in individual parameters of voice quality postintervention	No significant difference between control and intervention group on speech, HRQOL and self-management. Experimental group showed improved swallow function and communication sub-score	There was a subjective perceptual improvement in degree of dysphonia, roughness, breathiness; legato voice and pitch range. However this was not statistically tested

Abbreviations: SCL-90, symptom checklist-90-revised; O-C, obsession-compulsion; DEP, depression; PARA, paranoia; ANX, anxiety; HOS, hostility; WHO-QOL, World Health Organisation Quality of Life Scale; REL, social relationships; PHY, physical health; I-SECEL, Italian version of Self-Evaluation of Communication Experiences after Laryngeal Cancer; INFVo–I, overall impression; N, additive noise; F, fluency; Vo, voicing; VHI, Voice Handicap Index; MEP, maximum expiratory pressure; PEF, peak expiratory flow; VC, vital capacity; FEV1, forced expiratory volume in 1 s; COPD, chronic obstructive pulmonary disease; CPET, cardio pulmonary exercise testing; MPT, maximum phonation time; dB, decibels; EORTC-QLQ, European organisation for research and treatment of cancer quality of life questionnaire; PAM, patient activation measure; Swal-QOL, swallowing quality of life questionnaire; SHI, speech handicap index.

studies. Three studies[31,34,35] included breathing exercises. Two studies[31,34] provided practice of coordinating stoma occlusion, and three[31,32,34] included speech-related elements such as articulation or prosody exercises. Jansen et al, 2020[32] was the only study to include upper body flexibility and lymphoedema exercises. All included studies varied in the dosage and intensity of intervention.

Two studies[32,34] did not provide a rationale for the trialled interventions. Longobardi et al (2019)[31] provided a rationale for the investigation of psychological approaches, which was the main focus of the study, whereas rationale was not provided for the included voice production approaches. Similarly, Van Sluis et al (2020)[35] described clearly the rationale for targeting expiratory pressures through expiratory muscle strength training (EMST), however, the potential impact of EMST on alaryngeal voice was not elaborated.

Whilst the theoretical basis of the interventions was not illustrated in the studies, all interventions could be mapped onto the Van Stan Taxonomy of Voice Therapy.[23] However, it has not yet been established that the theoretical principles underpinning interventions for laryngeal voice can be transferred to alaryngeal voice therapy without modification, therefore, it is uncertain whether targeted outcomes could be influenced by the interventions.

Synthesis of findings and exploration of relationships within the data

The heterogeneity of outcome measures prevented combined analysis of findings, therefore different intervention targets are described narratively. In some studies outcome measures used did not directly align to the intervention. For example, breathwork was reported in Longobardi et al (2019)[31] and De Oliveira (2005)[34]; however, potential change was assessed using a measure of vocal function, rather than a specific breathing outcome measure.

Self-reported voice impact. The SECEL, VHI-10 and SHI were used to assess self-reported voice impairment in Longobardi et al, (2019);[31] Van Sluis et al (2020)[35] and Jansen et al (2020)[32] respectively. Jansen et al (2020)[32] also used the SWAL-QOL, which includes a subsection score for communication. SECEL scores improved over time with a significant difference in pre postintervention mean scores ($P < 0.001$), representing reduced self-reported voice impairment with a large effect size ($d = 1.61$) following intervention in Longobardi et al (2019).[31] SECEL scores improved in both control and experimental groups. This was consistent with both groups receiving the same communication-related interventions, as the independent variable of the study was psychological support. Interaction effect was not reported; however, Jansen et al (2020)[32] reported significant improvement in the communication sub-score of the SWAL-QOL ($P = 0.004$) with small to medium effect size at postintervention and follow up ($d = -0.18, -0.53$) in the experimental group.

In contrast, no significant differences were found on the VHI-10 and SHI measures. This difference may be accounted for by variation in outcome measures. Significant change was demonstrated on the SECEL, which is a measure designed specifically for laryngeal cancer and validated within the laryngectomy population, whereas the VHI-10 is not designed for oncological populations. The SHI was developed and validated within an oropharyngeal cancer population, however, the focus of the scale is on speech problems, rather than voice quality, therefore it may not be the optimal assessment to measure alaryngeal voice change.

Studies that demonstrated significant change did not focus primarily on voice interventions. Jansen et al (2020)[32] and Longobardi et al (2019)[31] investigated the use of a self-help programme and psychological support respectively.

Vocal and pulmonary function. Longobardi et al (2019)[31] used the INFVo scale for the perceptual evaluation of voice quality. Significant improvement in prepost scores was reported in both experimental and control groups in all domains ($P < 0.022$) with large effect size ($d = -1.33$ to 5.80), and groups were significantly different ($P < 0.0005$), however, the interaction effect was not reported to enable interpretation of the finding. De Oliveira et al (2005)[34] did not find any significant differences postintervention in vocal quality, pitch or sentence melody, based on Likert scale perceptual ratings. Validated outcome measures were not used within the study.

Van Sluis et al (2020)[35] reported an increase in mean decibels (Db) postintervention (week four), which was not sustained at the second follow-up (week eight). This aligned to an increase in mean maximum expiratory pressure (MEP) at week four and slight mean decrease at week eight. Intensity, as measured in decibels, is proportional to expiratory pressure[41] therefore, the pattern of change in mean MEP and Db scores over time is consistent. Statistical tests were not used to determine significance, as the authors reported a lack of sufficient data on the laryngectomy population to perform sample size calculations. There was no change in mean maximum phonation time or pitch, which are not anticipated to be impacted by expiratory pressure changes. There was no change over time in other measures of pulmonary function, physical exertion, or fatigue.

Intelligibility. Intelligibility was measured within two studies[31,34] using perceptual rating. Whilst both studies included speech production or articulation exercises in their intervention programmes, only Longobardi et al (2019)[31] reported a significant improvement in intelligibility over time ($P < 0.001$). Data was not provided for calculation of effect size.

Whilst detail on the specific exercises was limited, this difference in outcome was unexpected as the speech

intervention described in De Oliveira et al (2005)[34] targeted speech production more directly through articulation exercises. This finding could be attributed to a difference in the measurement of intelligibility where Longobardi et al (2019)[31] used mean percentage of correctly perceived words, De Oliveira et al (2019)[34] used a perceptual Likert scale which may be less robust.

Psychological distress and quality of life. The two RCTs[31,32] included measures of psychological distress, well-being, and quality of life. Quality of life measures and outcomes differed between studies. Jansen et al (2020)[32] used the EORTC QLQ-C30 and H&N35 measures which did not find any significant group, time or interaction effects. These measures are specific to the general cancer and the head and neck cancer populations. Longobardi et al (2019)[31] assessed quality of life using the WHO-QOL, which is a generic measure of the quality of life, not specific to cancer populations. The study reported a different pattern of change in the social relationships domain where QOL improved in the experimental group and declined in the control group ($P = 0.017$) with a large effect size ($d = -1.14$); and over time only ($P < 0.004$) in the psychological health domain. Longobardi et al (2015)[31] additionally measured self-reported psychological distress. This demonstrated a significant effect of the group on depression ($P = 0.017$), paranoid ideation ($P = 0.038$), and obsessive-compulsiveness ($P = 0.013$) subscores with large effect sizes; where experimental group scores decreased and control group scores increased (indicative of more severe symptoms).

The difference in outcomes between the two studies could be attributed to the inclusion of psychological intervention within Longobardi et al (2015).[31] Whilst the experimental group in Jansen et al (2020)[32] received additional weekly coaching from a healthcare professional, this was not targeting psychological support.

Robustness of synthesis

PRISMA guidance was followed to ensure that the review process adhered to an established framework, with a search strategy designed to source all available evidence. The small number of studies and quality limitations meant limited data were available for synthesis.

DISCUSSION

This systematic review sought to identify and appraise therapeutic interventions for voice rehabilitation following total laryngectomy with tracheoesophageal voice prosthesis. The review searched eight healthcare databases and followed PRISMA guidance. Six studies were identified which met criteria for inclusion. This included two randomised controlled trials, one single case study and three before and after studies. All studies presented quality issues with two studies excluded from data synthesis due to a high risk of bias. The included studies explored voice rehabilitation after total laryngectomy, however, this was not the primary focus of the intervention in the majority of studies. Interventions varied between studies with inadequate description for replication. Evidence for the effectiveness of the interventions was limited with few significant findings and inconsistent outcomes between studies. Most studies contained small sample sizes and male participants were over-represented, therefore, the generalisability of the outcomes should be interpreted with caution. Furthermore, few participants had received extended surgical reconstructions therefore the results may not be generalised to participants outside of the total laryngectomy with primary closure population.

Interventions for tracheoesophageal voice

The TIDieR checklist[27] was used to assess intervention reporting. This demonstrated that interventions were not well described with insufficient detail for replication in some studies. This finding is consistent with the literature on laryngeal voice therapy interventions.[42] There was limited information on how specific exercises were carried out by participants and an absence of a theoretical rationale for some approaches. Where reported, studies varied in the number of sessions provided, intensity, and frequency of intervention. There was significant heterogeneity in choice of outcome measures with some reliance on measures designed outside of the laryngectomy population or informal measures such as Likert scales. This accords with systematic reviews of laryngeal voice therapy[43,44] which found heterogeneity and methodological issues with the laryngeal voice evidence base. In the laryngectomy context, variation between studies and scarce theoretical basis may point towards a lack of consensus or awareness of which interventions or outcome measures could be of benefit to voice prosthesis users.

Whilst interventions could be mapped onto the laryngeal voice therapy categorisations, there was an absence of direct 'borrowing' from established laryngeal voice therapy methods and certain aspects of voice were not addressed within the included studies. For example, studies did not include approaches for resonance, and pitch was incorporated only via singing exercises rather than speech. This contrasts with the established use of semi-occluded vocal tract exercises[45] and vocal function exercises[46] to target resonance and pitch in laryngeal voice therapy. Functional approaches to voice use were not seen in identified studies, yet this approach is well established in laryngeal voice therapy, for example via Lee Silverman Voice Training,[47] which focuses on impairment-based drills plus functional tasks to transfer therapy gains to real-world environments. This may suggest uncertainty about the effectiveness of these approaches within the laryngectomy population.

Quality of evidence and effectiveness of voice interventions

Two studies[33,36] were excluded from analysis due to the high risk of bias, whilst quality issues were present in all studies. Studies tended to deliver an intervention of several composite parts whilst using outcome measures that were not directly related to specific components. In the majority of studies the primary outcome measure was not specified. This produced a lack of clarity on what specific element of the intervention could be beneficial and confounded interpretation of findings.

A small number of positive findings were reported across studies, however, there was inconsistency in outcomes and a lack of commonalities relating to outcome measures and intervention details, which reduced the robustness of positive findings. Consequently, the existing evidence base is not sufficient to inform therapeutic intervention. This points toward the need to develop clinical population-specific voice measures; which has similarly been highlighted within the non-laryngectomy head and neck cancer population[48] and other emerging fields of voice-related SLT practice, such as subglottic stenosis[49] and above-cuff vocalisation.[50] The necessity for further research to enhance transferrable evidence is identified across areas of voice therapy.[42,43,51]

Limitations

This review aimed to identify evidence for therapeutic rehabilitation of tracheoesophageal voice and to critically appraise the quality of the evidence and effectiveness of interventions. The review demonstrated that there is a very small evidence base in this area, finding only six studies, which presented risk of bias. Due to the heterogeneity of interventions and outcomes used it was not possible to perform a meta-analysis. The review has therefore found limited information on what therapeutic interventions may be appropriate for tracheoesophageal voice. Moreover the small number of studies and identified risk of bias further reduce confidence in the findings and generalisability within clinical practice.

CONCLUSION

This systematic review sought to find evidence that could inform clinical practice; however, this was lacking, and robust evidence was not found. This has implications for future research and highlights the need for more studies on the development and testing of therapeutic interventions to improve tracheoesophageal voice. Intervention development work should align to the Medical Research Council Framework for Developing and Evaluating Complex Interventions[52] and follow a dynamic process of development, testing, refinement with stakeholder engagement.

In acknowledging the requirement for further research, it is recommended that studies adhere to reporting guidelines such as TIDieR,[27] and methodological issues are considered to enhance quality. Consistent use of outcome measures that are designed and validated for the laryngectomy or head and neck cancer populations is advocated, to enable comparison of findings across studies and to improve robustness of outcomes. Further studies are required which explore multidimensional aspects of alaryngeal voice production such as pitch, resonance, and functional voice use; alongside wider investigation of the effectiveness of breathing and articulation approaches seen in the included studies. No studies considered the views of service users in detail within the development, study design, or intervention content and this is identified of an area of research importance.

Declaration of Competing Interest

The authors have no declarations of interest to disclose.

Appendix A. Supporting information

Supplementary data associated with this article can be found in the online version at doi:10.1016/j.jvoice.2023.10.033.

References

1. Kazi R. Surgical voice restoration following total laryngectomy. *J Cancer Res Ther.* 2007;3:188–189. https://doi.org/10.4103/0973-1482.38991.
2. Cinamon U, Levy D, Sokolov M, et al. Total Laryngectomy as a Palliative procedure for a patient with advanced laryngeal and colon cancers: dilemmas. *J Palliative Care.* 2006;22:309–311. https://doi.org/10.1177/082585970602200411.
3. McCammon SD. Concurrent palliative care in the surgical management of head and neck cancer. *J Surg Oncol.* 2019;120:78–84. https://doi.org/10.1002/jso.25452.
4. Cancer Research UK. (2022). Survival for laryngeal cancer. Available at: https://www.cancerresearchuk.org/about-cancer/laryngeal-cancer/survival. Accessed June 18, 2022.
5. Macmillan Cancer Support, 2017. The rich picture: People with head and neck cancer. Understanding the numbers, needs and experiences of people affected by cancer. Macmillan Cancer Support.
6. Clarke P, Radford K, Coffey M, et al. Speech and swallow rehabilitation in head and neck cancer: United Kingdom National Multidisciplinary Guidelines. *J Laryngol Otol.* 2016;130:S176–S180. https://doi.org/10.1017/S0022215116000608.
7. Schache A, Kerawala C, Ahmed O, et al. British association of head and neck oncologists (BAHNO) standards 2020. *J Oral Pathol Med.* 2021;50:262–273. https://doi.org/10.1111/jop.13161.
8. Evans, E. Hurren, A. Govender, R.2010. Prosthetic Surgical Voice Restoration (SVR): the role of the speech and language therapist. RCSLT Policy Statement.
9. Sharpe G, Camoes Costa V, Doube W, et al. Communication changes with laryngectomy and impact on quality of life: a review. *Qual Life Res.* 2019;28:863–877.
10. Perry A, Casey E, Cotton S. Quality of life after total laryngectomy: functioning, psychological well-being and self-efficacy. *Int J Lang Commun Disord.* 2015;50:467–475. https://doi.org/10.1111/1460-6984.12148.
11. Kemps GJF, Krebbers I, Walmari P, et al. Affective symptoms and swallow-specific quality of life in total laryngectomy patients. *Head and Neck.* 2020;42:3179–3187. https://doi.org/10.1002/hed.26365.

12. Keszte J, Danker H, Dietz A, et al. Mental disorders and psychosocial support during the first year after total laryngectomy: a prospective cohort study. *Clin Otolaryngol.* 2013;38:494–501. https://doi.org/10.1111/coa.12194.
13. Kam D, Salib A, Gorgy G. Incidence of suicide in patients with head and neck cancer. *JAMA Otolaryngol Head Neck Surg.* 2015;141:1075–1081. https://doi.org/10.1001/jamaoto.2015.2480.
14. Luu K, Chang BA, Valenzuela D, et al. Primary versus secondary tracheoesophageal puncture for voice rehabilitation in laryngectomy patients: a systematic review. *Clin Otolaryngol.* 2018;43:1250–1259. https://doi.org/10.1111/coa.13138.
15. Mahalingam S, Srinivasan R, Spielmann P. Quality of life and functional outcomes following pharyngolaryngectomy: a systematic review of literature. *Clin Otolaryngol.* 2016;41:25–43. https://doi.org/10.1111/coa.12466.
16. Pawar PV, Sayed SI, Kazi R, et al. Current status and future prospects in prosthetic voice rehabilitation following laryngectomy. *J Cancer Res Ther.* 2008;4:186–191.
17. Shuxin X. Effectiveness of voice rehabilitation on vocalisation in post-laryngectomy patients: a systematic review. *Int J Evid-Based Healthc.* 2010;8:256–258.
18. Van Sluis KE, der Molen L, Van Son RJJH, et al. Objective and subjective voice outcomes after total laryngectomy: a systematic review. *Eur Arch Oto-Rhino-Laryngol.* 2018;275:11–26. https://doi.org/10.1007/s00405-017-4790-6.
19. Singer S, Merback M, Dietz A, et al. Psychosocial determinants of successful voice rehabilitation after laryngectomy. *J Chin Med Assoc.* 2007;70:407–423.
20. Coffey MM, Tolley N, Howard D, et al. Evaluating the effect of different voice prostheses on alaryngeal voice quality. *Laryngoscope.* 2018;128:2460–2466.
21. Tawfik GM, Makram OM, Zayan AH, et al. Voice rehabilitation by voice prosthesis after total laryngectomy: a systematic review and network meta-analysis for 11,918 patients. *J Speech Lang Hear Res.* 2021;64:2668–2681.
22. Page MJ, McKenzie JE, Bossuyt PM, et al. The PRISMA 2020 statement: an updated guideline for reporting systematic reviews. *Br Med J.* 2021;372:N71. https://doi.org/10.1136/bmj.n71.
23. Van Stan JH, Roy N, Awan S, et al. A taxonomy of voice therapy. *Am J Speech-Lang Pathol.* 2015;24:101–125.
24. National Heart, Lung and Blood Institute (NLHBI). Quality assessment tool for before-after (pre-post) studies with no control group. *Study Qual Assess Tools.* 2021.⟨https://www.nhlbi.nih.gov/health-topics/study-quality-assessment-tools⟩.
25. Maher CG, Sherrington C, Herbert RD, et al. Reliability of the PEDro scale for rating quality of randomised control trials. *Phys Ther.* 2003;83:713–721.
26. Tate RL, Perdices M, Rosenkoette U, et al. Revision of a method quality rating scale for single case experimental designs and n-of-1 trials: the 15-item risk of bias in N-of-1 trials (RoBiNT) scale. *Neuropsychol Rehabil.* 2013;23:619–638.
27. Hoffmann TC, Glasziou PP, Milne R, et al. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *Br Med J.* 2014;348:g1687.
28. Cohen J. *Statistical Power Analysis for the Behavioural Sciences*. revised ed. New York: Academic Press; 1977.
29. Lakens D. Calculating and reporting effect sizes to facilitate cumulative science: a practical primer for t-tests and ANOVAS. *Front Psychol.* 2013;4(863):1–12. https://doi.org/10.3389/fpsyg.2013.00863.
30. Popay J, Roberts H, Sowden A, et al. Guidance on the conduct of narrative synthesis in systematic reviews. *ESRC Methods Programme.* 2006:11–87.Version 1.
31. Longobardi Y, Savoia V, Bussu F, et al. Integrated rehabilitation after total laryngectomy: a pilot trial study. *Support Care Cancer.* 2019;27:3537–3544. https://doi.org/10.1007/s00520-019-4647-1.
32. Jansen F, Cnossen IC, et al. Effectiveness of a guided self-help exercise program tailored to patients treated with total laryngectomy: results of a multi-center randomised controlled-trial. *Oral Oncol.* 2020;103(104586) https://doi.org/10.1016/j.oraloncology.2020.104586.
33. Onofre F, Ricz HMA, Takeshita-Monaretti TK, et al. Effect of singing training on total laryngectomees wearing a tracheoesophageal voice prosthesis. *Acta Cirugica Bras.* 2013;28:119–125.
34. De Oliveira IB, Costa CC, Cagas JF, et al. Oral communication of laryngectomised patients with tracheoesophageal prosthesis. Pre and post-treatment comparative analysis. *Pro-fono Rev Atualizacao Cientifica.* 2005;17:165–174.
35. Van Sluis KE, Kornman AF, Groen WG, et al. Expiratory muscle strength training in patients after total laryngectomy: a feasibility pilot study. *Ann Otolaryngol Rhinol Laryngol.* 2020;129:1186–1194.
36. Baijens LWJ, Speyer R, Roodenburg N, et al. Rehabilitation program for prosthetic tracheojejunal voice production and swallowing function following circumferential pharyngolaryngectomy and neopharyngeal reconstruction with a jejunal free flap. *Dysphagia.* 2010;26:78–84.
37. Jansen F, Cnossen IC, Eerenstein SEJ, et al. Effectiveness and cost-utility of a guided self-help exercise program for patients treated with total laryngectomy: protocol of a multi-centre randomised controlled trial. *BMC Cancer.* 2016;16:580. https://doi.org/10.1186/s12885-016-2613-6.
38. Hirano M. *Clinical Examination of Voice*. Vienna: Springer-Verlag publishing; 1981.
39. Jacobsen BH, Johnson A, Grywalski C, et al. The voice handicap index (VHI): development and validation. *Am J Speech-Lang Pathol.* 1997;6:66–70.
40. Rinkel RN, Verdonck-de Leeuw IM, Van Reij EJ, et al. Speech Handicap Index in patients with oral and pharyngeal cancer: better understanding of patients' complaints. *Head Neck.* 2008;30:868–874.DOI: 10/1002/hed.20795.
41. Huber JE, Chandrasekaran B, Wolstencroft JJ. Changes to respiratory mechanisms during speech as a result of different cues to increase loudness. *J Appl Physiol.* 2005;98:2177–2184. https://doi.org/10.1152/japplphysiol.01239.2004.
42. White AC, Awad R, Carding P. Pre and post-operative voice therapy intervention for benign vocal fold lesions: a systematic review. *J Voice.* 2021;37(6):857–874.
43. Desjardins M, Halstead L, Cooke M, et al. A systematic review of voice therapy: what 'effectiveness' really implies. *J Voice.* 2016;31:p13–p32.
44. Speyer R. Effects of voice therapy: a systematic review. *J Voice.* 2006;22:565–580.
45. Titze IR. Voice training and therapy with a semi-occluded vocal tract: rationale and scientific underpinnings. *J Speech Lang Hear Res.* 2006;49:448–459.
46. Stemple JC, Lee L, Damico B, et al. Efficacy of vocal function exercises as a method of improving voice production. *J Voice.* 1994;8:270–278.
47. Pannbacker M. Voice treatment techniques: a review and recommendations for outcome studies. *Am J Speech-Lang Pathol.* 1998;7:49–64.
48. Jacobi I, van der Molen L, Huiskens H, et al. Voice and speech outcomes of chemoradiation for advanced head and neck cancer: a systematic review. *Eur Arch Oto-Rhino-Laryngol.* 2010;267:1495–1505.
49. Clunie GM, Roe JWG, Alexander C, et al. Voice and swallowing outcomes following airway reconstruction in adults: a systematic review. *Laryngoscope.* 2020;131:p146–p157.
50. Mills CS, Michou E, King N, et al. Evidence for above cuff vocalisation in patients with a tracheostomy: a systematic review. *Laryngoscope.* 2022;132:p600–p611.
51. Ramachandran A, Hilari K, Epstein R, et al. Voice related coping in professional voice users: a systematic review. *J Voice.* 2022.
52. Skivington K, Matthews L, Simpson SA, et al. A new framework for developing and evaluating complex interventions: update of Medical Research Council guidance. *Br Med J.* 2021;374:n2061. https://doi.org/10.1136/bmj.n2061.