Voice outcomes post total laryngectomy

Anne Hurren and Nick Miller

Anne Hurren PhD
Speech and Language Sciences, Leeds Beckett University, PD405 City Campus, Leeds, UK LS1 3HE
Anne Hurren: a.hurren@leedsbeckett.ac.uk

Telephone: 07905 681769

Nick Miller PhD
Institute of Health and Society, Speech and Language Sciences, Newcastle University, Newcastle upon Tyne, UK NE1 7RU
Nick Miller: nicholas.miller@newcastle.ac.uk

Acknowledgements: None
Financial Support and Sponsorship: AH has undertaken consultancy work for PROTiP Medical and In Health Technologies and their UK distributor. Honoraria and sponsorship for training events have been received from In Health Technologies and their UK distributor.
Conflicts of Interest: NM has no conflicts of interest

Abstract

Purpose of review: A consensus and body of robust evidence has developed regarding optimal laryngeal voice outcome measures. This contrasts with a lack of clarity for equivalent assessments in alaryngeal voice. Addressing this situation would enable clinicians to select the best tools currently available to facilitate research, audit and clinical practice. This is important because of the limited knowledge regarding the optimal surgical or reconstruction techniques and rehabilitation regimes for the laryngectomy population.

Recent findings: There is currently no evidence to support the use of acoustic instrumental measures in terms of validity. Preliminary data supports the validity of a new tracheoesophageal voice auditory–perceptual tool the SToPS, for professional and naïve raters. Few specific self-rating tools exist with the SECEL having the most evidence regarding validity, reliability and clinical utility. Laryngeal self-report questionnaires have been utilised but concerns have been expressed regarding content validity. Patient self-report outcomes do not concur with professional or naïve judgements which reflects findings in the laryngeal voice literature.

Summary: Further research is needed to establish the optimal tools for research and clinical practice. Investigations should also incorporate assessments of real life communication in daily living rather than solely focussing on recordings in laboratory conditions.

Key words or phrases: laryngectomy; alaryngeal voice; tracheoesophageal voice; voice quality assessment; voice outcome measures.

Introduction

Outcome measurement for laryngeal voice has developed rapidly over the last 15 years. There now exists a considerable evidence base and consensus regarding which assessments to utilise in research and clinical practice. These have recently been summarised [1]** and has important implications for clinicians working in the field of laryngectomy; alaryngeal voice studies have frequently failed to incorporate applicable lessons learned from the more advanced research in laryngeal voice quality.

In contrast, there is no current agreement for the optimal assessment of alaryngeal voice. Most studies directed at clinical practice or as research evaluation have utilised informal tools or acoustic measures with limited consideration of feasibility, validity or clinical application and reliability. This paper offers a summary and critical appraisal of recent research, with the aim of facilitating clinicians' decision-making when selecting outcome measures in their research and/or clinical practice. Three types of assessment are considered: 1) perceptual voice scales, 2) self-rating scales/questionnaires and 3) acoustic and instrumental measures.

Alaryngeal perceptual voice scales

Despite important differences between laryngeal and alaryngeal voice, there is sufficient commonality to transfer some key methods and findings of laryngeal perceptual assessment into evaluation of laryngectomy voice quality. A recent state of the art summary concerning all aspects of laryngeal voice quality assessment, including instrumental tools, concluded that auditory perceptual judgement continues to be the gold standard [2]**. However, there are major issues in simply transferring well evidenced support for e.g. GRBAS and CAPE-V scales [3-5] to the laryngectomy population. Firstly, content validity issues apply, as such scales do not capture the key vibratory features of the new phonatory mechanism (the "neoglottis"). Secondly, given the fundamental changes to the phonatory mechanism it is not possible for alaryngeal speakers to ever achieve a rating of normal laryngeal voice [6] thus rendering them impractical and invalid for charting change, since only one part of the scales will ever apply.

In response to these issues three studies have aimed to design and evaluate a perceptual scale specifically for alaryngeal voice quality [6-9].

Van As et al introduced 7-point bipolar scales (e.g. a scale from 'pleasant' to 'unpleasant') plus an "Overall Voice Judgement" 3-point semantic scale that measures rater impression of overall voice quality as "good", "reasonable" or "poor". Despite the positive move to develop a tool specific to alaryngeal voice, the scales have limitations. They lack: a) readily discernible baselines against which to measure e.g. it would be difficult to ascertain a norm for "unpleasant-pleasant" even for laryngeal speakers; b) guidelines for how to grade voice stimuli for each bipolar parameter, e.g. what counts as a score of 2 versus 4 for "noise-no noise"; c) drawbacks in calculations for rater reliability. In particular their agreement data was calculated with co-efficients criticised for artificially inflating agreement and designed for internal scale consistency [10, 11]. Further, the tool has no follow-up data and the author herself utilised other scales in later research on perceptual judgements [12-14].

The INFVo scale is a five parameter visual analogue scale devised by Moerman et al [8] and subsequently investigated in Italian translation [15]. Its design specifies it can assess "substitution voice" of any vibratory source other than the vocal folds e.g. neoglottis in total laryngectomy or false vocal fold or other sources in partial laryngectomy. Whilst this facilitates outcome comparison between partial and total laryngectomy, it still has key omissions. It neglects content validity, as the

key factor in total laryngectomy phonation relates to neoglottal tone [16] but this feature cannot be measured by the scales. Moerman et al also failed to specify a baseline against which voices can be measured nor considered chance agreement calculation in evaluating rater reliability.

The SToPS [6, 7] aimed to address the limitations of previous scales by designing a unique tool to circumvent these drawbacks. "Overall Grade" in the SToPS is measured against an internalised baseline of the most optimal tracheoesophageal voice it is possible to achieve. Written guidelines specify how raters should judge each scale point for all 14 parameters. It controls for possible chance agreement when calculating reliability and takes content validity (systematic review and pilot studies) into account. Furthermore, this is the first study to investigate the effect of rater type (by profession; by experience) on agreement ratings for tracheoesophageal voice. Findings suggested that SLT's (speech-language therapists) with experience in both dysphonia and laryngectomy may attain higher agreement than SLTs with only a head and neck oncology background, assumedly due to more stable internalised representations of tracheoesophageal voice. In contrast, no higher agreement co-efficients occurred with experienced Ear, Nose and Throat (ENT) surgeons compared to less experienced colleagues.

The SToPS is also the first alaryngeal tool to demonstrate reliability in a subsequent study that also used co-efficients that account for chance agreement [17]. Coffey [17] selected the SToPS to investigate voice outcome according to tracheoesophageal prosthesis type on the basis of clinical relevance, ease of use, clear terminology and guidance notes. Coffey reported thirteen of the sixteen parameters attained "good" agreement [18]. Coffey's finding of non-optimum agreement for the parameter "Tonicity" contrasts with Hurren et al [6] who reported "good" agreement provided raters were SLTs and experienced in voice. Coffey [17] suggested her finding may relate to the more complex scale design for "Tonicity" whereas Hurren attributed the superior agreement achieved solely by experienced SLTs to acquisition of internal representations of neoglottal tone needing more auditory-perceptual expertise. These discrepancies require more research.

The studies discussed above relate to professional judges, but naïve rater perspectives must also be considered because they represent the members of the community that patients will encounter in daily life. Naïve raters were previously considered unable to use scales to rate dysphonic voices [19, 20]. However, a recent study demonstrated they could reliably rate "Overall Grade" of tracheoesophageal voice, attaining "good" inter rater agreement with SLT and ENT judges [7] and good test-retest results. This is of clinical relevance, as it indicates that it is possible for the MDT to utilise measures of overall voice quality that reflect that of the community in which the person lives and can thus be incorporated into rehabilitation outcome measurement perspectives.

Alternative measures to equally appearing interval and visual analogue scale formats are also available. Direct magnitude estimation (DME) has been utilised to investigate naïve raters' perceptions of laryngeal and tracheoesophageal voice [19, 21, 22]. Here judges assign a score in relation to an agreed baseline stimulus (modulus) of zero. This provides an anchor comparator baseline and is argued to allow more sensitivity to change. DME has been considered not ready for clinical use due to issues in providing consensus moduli [22]. However, the new finding of considerable agreement for Overall Grade [7] opens up this possibility. Another technique, NeAR, is a rank and sort software programme [23] that permits voice stimuli to be arranged in severity, without having to assign limited categories on scales. This would allow one to chart changes in individual parameters in relation to each other.

The consistent inclusion of intelligibility in alaryngeal, but not laryngeal, assessment scales warrants more detailed discussion. Given that rehabilitation always aims to achieve improved and optimum

intelligibility, then lack of intelligibility testing would represent a serious research and clinical omission. However restricting assessment to clinical settings or audio recordings in research type conditions is problematic in relation to reduced intelligibility being linked to speaking over background noise or other real life situations which may be an especial issue for alaryngeal speakers [24]**. Intelligibility fluctuations have also been linked to the variable ability of the neoglottis to signal voiced/voiceless distinctions [25, 26]. For a comprehensive review of issues in intelligibility assessment clinicians should refer to the summary by Miller [27]. A further variable in quantifying intelligibility concerns listener effort, which aims to capture the extra burden of cognitive processing load for listeners interpreting altered speech and voice [28, 29]. Intelligibility is an area warranting more detailed study to ensure the range of tracheoesophageal outcomes represent realistic, rather than clinical communication situations.

Self-rating scales/questionnaires

This aspect of voice evaluation is relatively well established in the field of laryngeal voice, with a considerable body of supporting evidence for validity and reliability [1]. Many informal, non-validated scales have been developed for research into alaryngeal voice e.g. [30-35] but there are limited investigations that specifically aim to develop, validate and assess the sensitivity of such tools in clinical practice. This represents a considerable deficit as patient perspectives are an essential component of multidimensional outcome measurement. In the absence of specific tools, assessments commonly used in laryngeal voice have been utilised. The Vocal Handicap Index (VHI) [36] and VHI-10 [37] show mild-moderate handicap in alaryngeal speakers [38-41]and the VoiSS [42] was sufficiently sensitive to demonstrate poorer outcomes for salvage compared to primary laryngectomy. The Voice Related Quality of Life (VRQOL) has been used in a number of alaryngeal investigations [40, 43-47] and has the added advantage of having undergone an investigation of construct validity for alaryngeal voice [46]. This involved factor analysis [46] which suggested that alaryngeal speakers interpret some questions from an awareness of the social impact of their voice due to its difference to that of laryngeal speakers. The authors considered the tool's validity with this patient group would be increased if these questions were rephrased or moved from the physical to the social-emotional domain section.

Generic communication tools also offer further outcome measurement possibilities. The Communicative Participation Item Bank (CPIM) ([24]** is a significant addition due to its validation with head and neck oncology patients plus allowing the previously discussed communication ability in everyday settings rather than a sole focus on voice related symptoms [48, 49] . Whilst the aforementioned scales allow research that compares outcomes of total laryngectomy compared to alternative larynx conservation options, limitations are likely to occur when a specific, sensitive tool with appropriate content validity is required. The Self Evaluation of Communication Experiences after Laryngectomy (SECEL) [50] assesses adjustment to communication and is the only such specific tool that has undergone validity and reliability testing and includes a composite overall score. It is unclear why the SECEL has not been widely adopted in clinical practice or research. Further studies relate to its adaptation to permit assessment of all laryngeal cancer treatments [51-53], though this currently remains available only in Swedish. The original SECEL has also been adapted into Italian [54] and Turkish [55].

Coffey's [17] investigation of voice prosthesis type in relation to voice and swallow outcomes highlighted the lack of a suitable tool to assess this for clinical purposes. This prompted Coffey [17] to develop a new tool to cover such issues as stoma occlusion and wet voice quality. The Voice Prosthesis Questionnaire (VPQ) [56] which covers items in addition to voice prosthesis/voice e.g. humidification filters was judged to be impractically long. Coffey's new tool (Patient Communication and Swallowing with Voice Prosthesis Questionnaire) evolved from a consultation with twenty patients, thus addressing content validity. However, test-re-test reliability was not assessed. Similarly, to date, the SToPS patient scale has undergone inter-rater evaluation but not intra-rater agreement testing [7].

Although patient self-report is a fundamental part of clinical outcome measurement, the relationship of self-judgments to ratings by other groups has undergone limited investigation [24**, 57**, 58]. A key finding is low concurrence of patient judgements with those of other raters. Hurren et al [57]** found limited agreement between patients and their carer and varied agreement when these ratings were compared with naïve listener, SLT and ENT judgements, with variance being random rather than systematic. These findings are not surprising as evidence from laryngeal voice similarly reports that clinicians' and carers' perceptual ratings differ from patients' perceptions [59]. Similarly, naïve rater judgement of intelligibility in a quiet setting did not predict self-report with the VHI-10 or CPIB [24]**. However, when patients were intelligible over background noise there was a strong relationship with the self-report measures. It is possible that the lack of agreement relates to the artefact of patients experiencing real life communication situations whilst other raters judge from laboratory style recordings. Fruitful areas for future research would be to investigate auditory-perceptual assessments in real life situations.

Acoustic Measures

Many measurements come under the umbrella of acoustic assessment and are ubiquitous in both laryngeal [2]** and alaryngeal voice studies. Jitter, shimmer, long-term average spectra, harmonics to noise ratio and signal to noise ratio are most frequently selected as laryngeal voice outcome [1]**. However there are concerns that suggest such measures cannot be considered sufficiently robust for research and clinical practice [1, 60, 61]. These relate to: a) validity when only vowels are assessed; b) there are question marks over the feasibility/ appropriateness of accurate quantification of voice when there is marked aperiodicity; c) lack of evidence regarding inter and intra-user agreement; d) issues around comparability with auditory-perceptual measures and e) utility, as the procedure requires specific expertise and time (about one hour per patient). It is currently unclear how these acoustic measures can inform clinical practice regarding patient satisfaction or efficacy of treatment regimes.

Barsties and de Bodt [2]** comprehensively reviewed objective-acoustic analysis to include more recent studies of the Acoustic Voice Quality Index (AVQI) [62]* and the Cepstral Spectral Index of Dysphonia (CSID) [63]* that can assess both vowels and running speech. The CSID has demonstrated potential to distinguish normal from dysphonic voice [63] and severity of the overall grade of dysphonia [64] in comparison to auditory-perceptual judgements although there is limited correlation with the VHI [65]. The AVQI has shown ability to determine severity of overall grade [62]

and sensitivity to change [66]. Although these new tools have not yet been incorporated into tracheoesophageal voice research they currently have limited clinical application and it appears to be more ecologically valid to use perceptual rating scales to assess laryngeal and alaryngeal voice quality outcome.

Acoustic signal typing, previously utilised in laryngeal voice [67], has been developed for tracheoesophageal voice [12, 68]. This involved categorising tracheoesophageal voices into four sub-types based on raters' visual assessment of the acoustic content of narrow-band spectrograms supported by written guidelines [68]. Although a significant relationship was found between signal types and auditory-perceptual voice assessment for overall grade [68], a later study [12] concluded it offers " limited predictive information on voice quality" and cannot differentiate the key alaryngeal voice determinant of tonicity i.e. hypotonicity versus hypertonicity. Further limitations relate to only two judges being used to assess inter reliability and intra rater agreement was not included. Consequently this tool has yet to establish its validity and reliability and there is no evidence it can surpass or enhance auditory-perceptual tools as the optimal outcome measure.

Conclusion:

There remains a lack of consensus and limited evidence regarding the optimal tools to use for laryngectomy voice outcome, whether this is charting change acoustically or clinicians or naïve raters conducting perceptual assessment or patients self-rating their impressions. Perceptual voice rating scales show similar findings to laryngeal measurement scales and there is some preliminary evidence that there is sufficient reliability for the SToPS; Overall Grade is the most reliable measure and naïve raters have used the scale reliably and in agreement with professional judges. Laryngeal self-rating scales have been transferred to use with laryngectomy voice but there are some concerns regarding their validity. Acoustic measures are considered problematic in laryngeal voice and whilst new tools are developing there is currently no indication that such outcomes are sufficiently robust or reliably informative to use for alaryngeal voice outcome measurement. In the meantime it would appear optimal to use the validated assessments for perceptual and patient self-report. Future research needs to clarify the reliability of alaryngeal auditory-perceptual scales, the clinical applicability and reliability of the SECEL or develop a suitable alternative and move to assessment in real life communication situations rather than laboratory conditions.

Key points:

- Perceptual voice assessments have the most evidence to support clinical and research use
- A new perceptual assessment, SToPS, has preliminary validity and reliability data to support use with professional and naïve raters
- There is a need to develop more specific tools for self-rating that are valid and reliable
- Patient self-report does not concur with professional or naïve judgements

References

Papers of particular interest, published within the annual period of review, have been highlighted as:
* of special interest
** of outstanding interest

1. ** Carding, P., *Evaluating the Effectiveness of Voice Therapy*. 2017, Compton: Oxford.
Chapter 8 summarises methods of assessing laryngeal voice outcome and provides tables of studies in relation to measures selected.
2. ** Barsties, B. and M. De Bodt, *Assessment of voice quality: Current state-of-the-art.* Auris Nasus Larynx, 2015. **42**(3): p. 183-188.
Summarises the current state of the art for laryngeal voice measurements including auditory perceptual judgement, acoustic analysis and aerodynamic measurement.
3. Kempster, G.B., Gerratt, B. R., Abbott, K. V. et al., *Consensus Auditory-Perceptual Evaluation of Voice: Development of a Standardized Clinical Protocol.* American Journal of Speech-Language Pathology, 2009. **18**(2): p. 124-132.
4. Nemr, K., Simoes-Zenari, M., Cordeiro, G. F. et al., *GRBAS and Cape-V Scales: High Reliability and Consensus When Applied at Different Times.* Journal of Voice, 2012. **26**(6): p. 6.
5. Webb, A.L., Carding, P. N., Deary, I. J.et al., *The reliability of three perceptual evaluation scales for dysphonia.* European Archives of Oto-Rhino-Laryngology, 2004. **261**(8): p. 429-434.
6. Hurren, A., A.J. Hildreth, and P.N. Carding, *Can we perceptually rate alaryngeal voice? Developing the Sunderland Tracheoesophageal Voice Perceptual Scale.* Clinical Otolaryngology, 2009. **34**(6): p. 533-538.
7. Hurren, A., *The development of a new rating scale for the perceptual assessment of tracheoesophageal voice quality following total laryngectomy*, PhD Thesis, *Institute of Health and Society*. 2014, Newcastle University, UK.
8. Moerman, M., Martens, J. P., Crevier-Buchman, L. et al., *The INFVo perceptual rating scale for substitution voicing: development and reliability.* European Archives of Oto-Rhino-Laryngology, 2006. **263**(5): p. 435-439.
9. van As, C.J., Koopmans-Van Beinum, F. J., Pols, L. C. W. et al., *Perceptual evaluation of tracheoesophageal speech by naive and experienced judges through the use of semantic differential scales.* Journal of Speech Language and Hearing Research, 2003. **46**(4): p. 947-959.
10. Gerratt, B.R. and J. Kreiman, *Theoretical and methodological development in the study of pathological voice quality.* Journal of Phonetics, 2000. **28**(3): p. 335-342.
11. McDowell, I. And C. Newell, *Measuring Health: A guide to rating scales and questionnaires*. 2006, New York: Oxford University Press.
12. Clapham, R.P., Van As-Brooks, C. J., Van Son, R. et al., *The Relationship Between Acoustic Signal Typing and Perceptual Evaluation of Tracheoesophageal Voice Quality for Sustained Vowels.* Journal of Voice, 2015. **29**(4): p. 7.
13. Ward, E.C., Hancock, K., Lawson, N. et al., *Perceptual Characteristics Of Tracheoesophageal Speech Production Using The New Indwelling Provox Vega Voice Prosthesis: A Randomized Controlled Crossover Trial.* Head and Neck-Journal for the Sciences and Specialties of the Head and Neck, 2011. **33**(1): p. 13-19.
14. Hancock, K., Ward, E., Lawson, N. et al., *A prospective, randomized comparative study of patient perceptions and preferences of two types of indwelling voice prostheses.* International Journal of Language & Communication Disorders, 2012. **47**(3): p. 300-309.
15. Schindler, A., Ginocchio, D., Atac, M.et al., *Reliability of the Italian INFVo scale and correlations with objective measures and VHI scores.* Acta Otorhinolaryngologica Italica: Organo Ufficiale Della Società Italiana Di Otorinolaringologia E Chirurgia Cervico-Facciale, 2013. **33**(2): p. 121-127.

16. Cheesman, A.D., Knight, J., Mcivor, J. et al., *Tracheo-oesophageal 'puncture speech'. An assessment technique for failed oesophageal speakers.* Journal of Laryngology & Otology, 1986. **100**(2): p. 191-9.
17. Coffey, M., *A comparison of Fiberoptic Endoscopic Evaluation and Videofluoroscopy in post laryngectomy swallowing, and swallow and voice evaluation with different voice prostheses.*, PhD Thesis, *Department of Medicine*. 2013, Imperial College London.
18. Landis, J.R. and G.G. Koch, *The measurement of observer agreement for categorical data.* Biometrics, 1977. **33**: p. 159-174.
19. Eadie, T.L. and P.C. Doyle, *Direct magnitude estimation and interval scaling of naturalness and severity in tracheoesophageal (TE) speakers.* Journal of Speech Language and Hearing Research, 2002. **45**(6): p. 1088-1096.
20. Kreiman, J., B.R. Gerratt, and G.S. Berke, *The Multidimensional Nature of Pathological Vocal Quality.* Journal of the Acoustical Society of America, 1994. **96**(3): p. 1291-1302.
21. Eadie, T.L. and P.C. Doyle, *Auditory-perceptual scaling and quality of life in tracheoesophageal speakers.* Laryngoscope, 2004. **114**: p. 753-759.
22. Eadie, T.L. and P.C. Doyle, *Scaling of voice pleasantness and acceptability in tracheoesophageal speakers.* Journal of Voice, 2005. **19**(3): p. 373-383.
23. Gould, J., Waugh, J., Carding, P. et al., *A new voice rating tool for clinical practice.* Journal of Voice, 2012. **26**(4): p. e163-e170.
24.** Eadie, T.L., Otero, D. S., Bolt, S. et al., *The Effect of Noise on Relationships Between Speech Intelligibility and Self-Reported Communication Measures in Tracheoesophageal Speakers.* American Journal of Speech-Language Pathology, 2016. **25**(3): p. 393-407.
Investigation of naïve rater perceptions of intelligibility in quiet and simulated background noise situations in relation to VHI-10 and CPIB self rating. Speakers with better intelligibility against noise were demonstrated to have lower degrees of voice handicap and increased communication participation.
25. Jongmans, P., Van As, C. J., Pols, L. C. W. et al. *An introduction to the assessment of intelligibility of tracheoesophageal speech*. in *I.F.A.* 2003. Institute of Phonetic Sciences, University of Amsterdam, Netherlands.
26. Saito, M., M. Kinishi, and M. Amatsu, *Acoustic analyses clarify voiced-voiceless distinction in tracheoesophageal speech.* Acta Oto-Laryngologica, 2000. **120**(6): p. 771-777.
27. Miller, N., *Measuring up speech intelligibility.* International Journal of Language & Communication Disorders, 2013. **48**(6): p. 601-612.
28. Nagle, K.F. and T.L. Eadie, *Listener effort for highly intelligible tracheoesophageal speech.* Journal of Communication Disorders, 2012. **45**(3): p. 235-245.
29. Rogerson, J. and B. Dodd, *Is there an effect of dysphonic teachers' voices on children's processing of spoken language.* Journal of Voice, 2005. **19**(1): p. 47-60.
30. Brown, D.H., Hilgers, F. J. M., Irish, J. C. et al., *Postlaryngectomy voice rehabilitation: State of the art at the millennium.* World Journal of Surgery, 2003. **27**(7): p. 824-831.
31. Clements, K.S., Rassekh, C. H., Seikaly, H et al., *Communication after laryngectomy - An assessment of patient satisfaction.* Archives of Otolaryngology-Head & Neck Surgery, 1997. **123**(5): p. 493-496.
32. Delsupehe, K., Zink, I., Lejaegere, M. et al., *Prospective randomized comparative study of tracheoesophageal voice prosthesis: Blom-Singer versus Provox.* Laryngoscope, 1998. **108**: p. 1561-1565.
33. Heaton, J.M., Sanderson, D., Dunsmore, I. R. et al., *Speech assessment of patients using three types of indwelling tracheo-oesophageal voice prostheses.* Journal of Laryngology and Otology, 1996. **110**(4): p. 343-347.
34. Vlantis, A.C., Gregor, R. T., Elliot, H.et al., *Conversion from a non-indwelling to a Provox (R) 2 indwelling voice prosthesis for speech rehabilitation: comparison of voice quality and patient preference.* Journal of Laryngology and Otology, 2003. **117**(10): p. 815-820.

35. Watson, J.B. and S.E. Williams, *Laryngectomees and Nonlaryngectomees Perceptions of 3 Methods of Alaryngeal Voicing.* Journal of Communication Disorders, 1987. **20**(4): p. 295-304.
36. Jacobson, B.H., Johnson, A., Grywalski, C., et al., *The Voice Handicap Index (VHI): development and validation.* American Journal of Speech-Language Pathology 1997. **6**(3): p. 66-70.
37. Rosen, C.A., Lee, A. S., Osborne, J., et al., *Development and validation of the Voice Handicap Index-10.* Laryngoscope, 2004. **114**(9): p. 1549-1556.
38. Azevedo, E.H.M., Montoni, N., Goncalves, J. et al., *Vocal Handicap and Quality of Life After Treatment of Advanced Squamous Carcinoma of the Larynx and/or Hypopharynx.* Journal of Voice, 2012. **26**(2): p. E63-E71.
39. Evans, E., P. Carding, and M. Drinnan, *The Voice Handicap Index with post-laryngectomy male voices.* International Journal of Language & Communication Disorders, 2009. **44**(5): p. 575-586.
40. Kazi, R., De Cordova, J., Singh, A. et al., *Voice-related Quality of Life in Laryngectomees: Assessment Using the VHI and V-RQOL Symptom Scales.* Journal of Voice, 2007. **21**(6): p. 728-734.
41. Moerman, M., J.P. Martens, and P. Dejonckere, *Application of the Voice Handicap Index in 45 patients with substitution voicing after total laryngectomy.* European Archives of Oto-Rhino-Laryngology, 2004. **261**(8): p. 423-428.
42. Robertson, S.M., Yeo, J. C. L., Dunnet, C., et al., *Voice, swallowing, and quality of life after total laryngectomy: results of the west of Scotland laryngectomy audit.* Head & Neck, 2012. **34**(1): p. 59-65.
43. Day, A.M.B. and P.C. Doyle, *Assessing Self-Reported Measures of Voice Disability in Tracheoesophageal Speakers.* Journal of Otolaryngology-Head & Neck Surgery, 2010. **39**(6): p. 762-768.
44. Moukarbel, R.V., Doyle, P. C., Yoo, J. H. et al., *Voice-Related Quality Of Life (V-Rqol) Outcomes In Laryngectomees.* Head and Neck-Journal for the Sciences and Specialties of the Head and Neck, 2011. **33**(1): p. 31-36.
45. Schindler, A., Mozzanica, F., Ginocchio, D. et al., *Voice-related quality of life in patients after total and partial laryngectomy.* Auris Nasus Larynx, 2012. **39**: p. 77-83.
46. Bornbaum, C.C., A.M.B. Day, and P.C. Doyle, *Examining the Construct Validity of the V-RQOL in Speakers Who Use Alaryngeal Voice.* American Journal of Speech-Language Pathology, 2014. **23**(2): p. 196-202.
47. Eadie, T.L. and B.C. Bowker, *Coping and Quality of Life After Total Laryngectomy.* Otolaryngology--Head and Neck Surgery, 2012. **146**(6): p. 959-965.
48. Eadie, T.L., Lamvik, K., Baylor, C. R. et al., *Communicative Participation and Quality of Life in Head and Neck Cancer.* Annals of Otology Rhinology and Laryngology, 2014. **123**(4): p. 257-264.
49. Baylor, C., Yorkston, K., Eadie, T. et al., *The Communicative Participation Item Bank (CPIB): Item Bank Calibration and Development of a Disorder-Generic Short Form.* Journal of Speech Language and Hearing Research, 2013. **56**(4): p. 1190-1208.
50. Blood, G.W., *Development and assessment of a scale addressing communication needs of patients with laryngectomies.* American Journal of Speech-Language Pathology, 1993. **2**: p. 82-90.
51. Finizia, C., B. Bergman, and J. Lindstrom, *A cross-sectional validation study of self-evaluation of communication experiences after laryngeal cancer - A questionnaire for use in the voice rehabilitation of laryngeal cancer patients.* Acta Oncologica, 1999. **38**(5): p. 573-580.

52. Finizia, C., C. Palme, and B. Bergman, *A longitudinal study of the Swedish Self-Evaluation of Communication Experiences after Laryngeal Cancer questionnaire in patients treated for laryngeal cancer.* Acta Oncologica, 2002. **41**(3): p. 262-268.
53. Tuomi, L., Karlsson, T., Johansson, M. et al., *Health-related quality of life and voice following radiotherapy for laryngeal cancer - a comparison between glottic and supraglottic tumours.* Acta Oncologica, 2015. **54**(1): p. 73-79.
54. Schindler, A., Mozzanica, F., Brignoli, F. et al., *Reliability and validity of the Italian self-evaluation of communication experiences after laryngeal cancer questionnaire.* Head & Neck, 2013. **35**(11): p. 1606.
55. Evitts, P.M., Kasapoglu, F., Demirci, U. et al., *Communication adjustment of patients with a laryngectomy in Turkey: Analysis by type of surgery and mode of speech.* Psychology Health & Medicine, 2011. **16**(6): p. 650-660.
56. Kazi, R., Singh, A., De Cordova, J., et al., *Validation of a voice prosthesis questionnaire to assess valved speech and its related issues in patients following total laryngectomy.* Clinical Otolaryngology, 2006. **31**(5): p. 404-410.
57.** Hurren, A., Drinnan, M., Miller, N. et al., *A comparison of patient and carer perceptions of tracheoesophageal voice quality in relation to speech & language therapist, ENT surgeon and naïve raters*, Conference and poster presentation at the *British Association of Head and Neck Oncology Scientific Meeting*. 2016: London.

Presentation based on PhD thesis - found limited agreement between patients and their carer and low agreement when these ratings were compared with naïve listener, SLT and ENT judgements. Variance being random rather than systematic.

58. Eadie, T.L., Day, A. M. B., Sawin, D. E., et al., *Auditory-perceptual speech outcomes and quality of life after total laryngectomy.* Otolaryngology--Head And Neck Surgery: Official Journal Of American Academy Of Otolaryngology-Head And Neck Surgery, 2013. **148**(1): p. 82-88.
59. Karnell, M.P., Melton, S. D., Childes, J. M., et al., *Reliability of clinician-based (GRBAS and CAPEN) and patient-based (V-RQOL and IPVI) documentation of voice disorders.* Journal of Voice, 2007. **21**(5): p. 576-590.
60. Carding, P.N., Wilson, J. A., Mackenzie, K. et al., *Measuring voice outcomes: state of the science review.* Journal of Laryngology and Otology, 2009. **123**(8): p. 823-829.
61. Maryn, Y., Roy, N., De Bodt, M. et al., *Acoustic measurement of overall voice quality: A meta-analysis.* Journal of the Acoustical Society of America, 2009. **126**(5): p. 2619-2634.
62. Barsties, B. and Y. Maryn, *The improvement of internal consistency of the Acoustic Voice Quality Index.* American Journal of Otolaryngology, 2015. **36**(5): p. 647-656.
63. Awan, S.N., Roy, N., Zhang, D. et al., *Validation of the Cepstral Spectral Index of Dysphonia (CSID) as a Screening Tool for Voice Disorders: Development of Clinical Cutoff Scores.* Journal of Voice, 2016. **30**(2): p. 130-144.
64. Peterson, E.A., Roy, N., Awan, S. N., et al., *Toward Validation of the Cepstral Spectral Index of Dysphonia (CSID) as an Objective Treatment Outcomes Measure.* Journal of Voice, 2013. **27**(4): p. 401-410.
65. Awan, S.N., N. Roy, and S.M. Cohen, *Exploring the Relationship Between Spectral and Cepstral Measures of Voice and the Voice Handicap Index (VHI).* Journal of Voice, 2014. **28**(4): p. 430-439.
66. Maryn, Y., M. De Bodt, and N. Roy, *The Acoustic Voice Quality Index: Toward improved treatment outcomes assessment in voice disorders.* Journal of Communication Disorders, 2010. **43**(3): p. 161-174.
67. Titze, I.R., *Workshop on acoustic voice analysis: summary statement*. 1994, National Center for Voice and Speech: Iowa City, IA.

68. van As-Brooks, C.J., Koopmans-Van Beinum, F. J., Pols, L. C. W et al., *Acoustic signal typing for evaluation of voice quality in tracheoesophageal speech.* Journal of Voice, 2006. **20**(3): p. 355-368.

Acknowledgements
None

Financial support and sponsorship
None

Conflicts of interest
AH has undertaken consultancy work for PROTiP Medical and In Health Technologies and their UK distributor. Honoraria and sponsorship for training events have been received from In Health Technologies and their UK distributor.

NM has no conflicts of interest