Optimising outcome assessment of voice interventions, II: sensitivity to change of self-reported and observer-rated measures

Citation for published version:

Digital Object Identifier (DOI):
10.1017/S0022215107007839

Link:
Link to publication record in Edinburgh Research Explorer

Document Version:
Publisher's PDF, also known as Version of record

Published In:
The Journal of Laryngology & Otology (JLO)

Publisher Rights Statement:

General rights
Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

Take down policy
The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact openaccess@ed.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.
Optimising outcome assessment of voice interventions, II: sensitivity to change of self-reported and observer-rated measures

I N Steen, K MacKenzie*, P N Carding†, A Webb†, I J Deary‡, J A Wilson†

Abstract

Objectives: A wide range of well validated instruments is now available to assess voice quality and voice-related quality of life, but comparative studies of the responsiveness to change of these measures are lacking. The aim of this study was to assess the responsiveness to change of a range of different measures, following voice therapy and surgery.

Design: Longitudinal, cohort comparison study.

Setting: Two UK voice clinics.

Participants: One hundred and forty-four patients referred for treatment of benign voice disorders, 90 undergoing voice therapy and 54 undergoing laryngeal microsurgery.

Main outcome measures: Three measures of self-reported voice quality (the vocal performance questionnaire, the voice handicap index and the voice symptom scale), plus the short form 36 (SF 36) general health status measure and the hospital anxiety and depression score. Perceptual, observer-rated analysis of voice quality was performed using the grade–roughness–breathiness–asthenia–strain scale. We compared the effect sizes (i.e. responsiveness to change) of the principal subscales of all measures before and after voice therapy or phonosurgery.

Results: All three self-reported voice measures had large effect sizes following either voice therapy or surgery. Outcomes were similar in both treatment groups. The effect sizes for the observer-rated grade–roughness–breathiness–asthenia–strain scale scores were smaller, although still moderate. The roughness subscale in particular showed little change after therapy or surgery. Only small effects were observed in general health and mood measures.

Conclusion: The results suggest that the use of a voice-specific questionnaire is essential for assessing the effectiveness of voice interventions. All three self-reported measures tested were capable of detecting change, and scores were highly correlated. On the basis of this evaluation of different measures' sensitivities to change, there is no strong evidence to favour either the vocal performance questionnaire, the voice handicap index or the voice symptom scale.

Key words: Voice; Voice Quality; Quality of Life; Outcome Assessment (Healthcare); Outcome Measures

Introduction

Dysphonia is a multifactorial disorder which affects expressive communication, mood and general health status. As discussed in the preceding paper, a number of well validated measures is now available to assess these different domains. The responsiveness to change of these very different tools is, however, much less clear. There are comparatively few studies of the outcomes of dysphonia treatment, and most use limited outcome data sets in a highly selected patient group (for example, vocal fold medialisation). Furthermore, the number of studies of surgical as opposed to voice therapy outcomes is exceedingly small. Our objective was to compare the sensitivity to change of a number of different self-reported, perceptual and quality of life measures following conservative (i.e. speech and language therapy) and surgical intervention in a large, heterogeneous group of voice patients.

The specific aims of the study were: (1) to estimate the responsiveness to change of self-reported and perceptual ratings of voice quality and voice-related quality of life, in a large cohort of heterogeneous...
dyshonous patients; and (2) to estimate the range of effect sizes for voice therapy and surgical interventions, using self-reported and observer-rated measures, in order to inform future prospective, controlled trials.

Methods
The approach adopted was to identify patients with voice-related problems and then, using a range of generic and voice-specific measures, to assess their quality of life before and after medical intervention.

Three self-reported voice scales were completed by patients. The vocal performance questionnaire8 consists of 12 items which address the physical aspects of the voice problem and its social and emotional impact, scored to give a total score. The voice handicap index4 is a 30-item questionnaire with questions grouped into three content domains, representing the functional, emotional and physical aspects of voice disorders. Its sensitivity to change in voice was evaluated on a sample of 37 subjects with various vocal fold abnormalities.5 The voice symptom scale6 is a 30-item scale with three content domains and a total score, the reliability and validity of which have been assessed in a series of studies involving over 800 subjects.7

Perceptual, observer-rated analysis of voice quality was performed using the grade–roughness–breathiness–asthenia–strain scale.8–10 All voices were recorded on digital audiotape, following a standard procedure, both before and after treatment.11 The recorded voice sample included rote counting and speaking the days of the week, prolonged /a/ and /i/ vowels, and three sentences from the Rainbow Passage. The five grade–roughness–breathiness–asthenia–strain parameters were scored using a four-point rating scale, from zero (normal) to three (extreme). Each participant was scored on a standard pro forma by an independent expert rater. Independent raters were blinded to treatment group and treatment status, but aware of each patient’s age and sex. The short form 36 (SF 36)12 is an extensively validated, self-administered, 36-item questionnaire assessing quality of life. It has eight subscales and two global domains (mental health and physical health), and a large body of normative data is available.13 The SF 36 is known to be abnormal in patients with voice disorders.14

All patients also completed the hospital anxiety and depression scale.15

Patients
One hundred and forty-four patients complaining of hoarseness and attending out-patient clinics in Newcastle and Glasgow were assessed by the above measures, before and after medical intervention. The patients included a subgroup of patients described in our companion paper.2 The patient exclusion criteria were: no intervention undertaken; defaulting from follow-up; laryngeal cancer; age less than 18 years; and impaired language or receptive communication skills. At the initial out-patient appointment, each participant completed the three self-reported voice questionnaires, the SF 36 and the hospital anxiety and depression scale, and also had their voice recorded. Ninety patients received a course of speech and language therapy, while 54 patients underwent laryngeal surgery.

Analysis
For each questionnaire or rating scale, the effect size was defined as the change in mean score divided by the standard deviation of change scores. The effect size is independent of scale and sample size and can be used to make comparisons between the different questionnaires and different groups of subjects. It is accepted16,17 that values around 0.2 represent small effect sizes, values around 0.5 represent medium effect sizes and values around 0.8 represent large effect sizes. If subjects experience an improvement in quality of life, the outcome measure with the largest effect size is clearly the most sensitive to change.

Results
Table 1 shows the mean baseline and follow-up scores for each patient group, along with the mean improvement in quality of life (with a 95 per cent confidence interval (CI)) and an estimate of effect size. A paired t-test indicates those improvements that are statistically significant.

Both groups of subjects reported medium to large improvements on all three voice questionnaires. The smallest changes were in the emotional subscale of the voice handicap index (effect sizes = 0.44 and 0.48 for speech and language therapy and surgery groups, respectively) and in the physical symptoms subscale of the voice symptom scale (effect sizes = 0.38 and 0.43 for speech and language therapy and surgery groups, respectively). The largest changes in the individual scales were in the physical aspects of voice subscale of the voice handicap index (effect sizes = 0.71 and 0.81 for speech and language therapy and surgery groups, respectively) and in the voice impairment subscale of the voice symptom scale (effect sizes = 0.78 and 1.00 for speech and language therapy and surgery groups, respectively). The effect sizes corresponding to the change in the total score, respective to speech and language therapy and surgery, for each of the three voice questionnaires, were: vocal performance questionnaire, 1.04 and 0.82; voice handicap index, 0.62 and 0.72; and voice symptom scale, 0.78 and 1.06. The two patient groups were very similar both at baseline and follow up (Figure 1), with no significant differences between them.

The changes in the three voice questionnaires were highly correlated, as follows: vocal performance questionnaire vs voice symptom scale, 0.74 (95 per cent CI: 0.65, 0.81); vocal performance questionnaire vs voice handicap index, 0.76 (95 per cent CI: 0.68, 0.83); and voice symptom scale vs voice handicap index, 0.83 (95 per cent CI: 0.76, 0.87). All differences were significant ($p < 0.0001$). The correlation between the voice symptom scale and the voice handicap index was greater, due in part to the small number of shared items between the two questionnaires. Changes in the subscale components of the voice


dyshonous patients; and (2) to estimate the range of effect sizes for voice therapy and surgical interventions, using self-reported and observer-rated measures, in order to inform future prospective, controlled trials.

Methods
The approach adopted was to identify patients with voice-related problems and then, using a range of generic and voice-specific measures, to assess their quality of life before and after medical intervention.

Three self-reported voice scales were completed by patients. The vocal performance questionnaire8 consists of 12 items which address the physical aspects of the voice problem and its social and emotional impact, scored to give a total score. The voice handicap index4 is a 30-item questionnaire with questions grouped into three content domains, representing the functional, emotional and physical aspects of voice disorders. Its sensitivity to change in voice was evaluated on a sample of 37 subjects with various vocal fold abnormalities.5 The voice symptom scale6 is a 30-item scale with three content domains and a total score, the reliability and validity of which have been assessed in a series of studies involving over 800 subjects.7

Perceptual, observer-rated analysis of voice quality was performed using the grade–roughness–breathiness–asthenia–strain scale.8–10 All voices were recorded on digital audiotape, following a standard procedure, both before and after treatment.11 The recorded voice sample included rote counting and speaking the days of the week, prolonged /a/ and /i/ vowels, and three sentences from the Rainbow Passage. The five grade–roughness–breathiness–asthenia–strain parameters were scored using a four-point rating scale, from zero (normal) to three (extreme). Each participant was scored on a standard pro forma by an independent expert rater. Independent raters were blinded to treatment group and treatment status, but aware of each patient’s age and sex. The short form 36 (SF 36)12 is an extensively validated, self-administered, 36-item questionnaire assessing quality of life. It has eight subscales and two global domains (mental health and physical health), and a large body of normative data is available.13 The SF 36 is known to be abnormal in patients with voice disorders.14

All patients also completed the hospital anxiety and depression scale.15

Patients
One hundred and forty-four patients complaining of hoarseness and attending out-patient clinics in Newcastle and Glasgow were assessed by the above measures, before and after medical intervention. The patients included a subgroup of patients described in our companion paper.2 The patient exclusion criteria were: no intervention undertaken; defaulting from follow-up; laryngeal cancer; age less than 18 years; and impaired language or receptive communication skills. At the initial out-patient appointment, each participant completed the three self-reported voice questionnaires, the SF 36 and the hospital anxiety and depression scale, and also had their voice recorded. Ninety patients received a course of speech and language therapy, while 54 patients underwent laryngeal surgery.

Analysis
For each questionnaire or rating scale, the effect size was defined as the change in mean score divided by the standard deviation of change scores. The effect size is independent of scale and sample size and can be used to make comparisons between the different questionnaires and different groups of subjects. It is accepted16,17 that values around 0.2 represent small effect sizes, values around 0.5 represent medium effect sizes and values around 0.8 represent large effect sizes. If subjects experience an improvement in quality of life, the outcome measure with the largest effect size is clearly the most sensitive to change.

Results
Table 1 shows the mean baseline and follow-up scores for each patient group, along with the mean improvement in quality of life (with a 95 per cent confidence interval (CI)) and an estimate of effect size. A paired t-test indicates those improvements that are statistically significant.

Both groups of subjects reported medium to large improvements on all three voice questionnaires. The smallest changes were in the emotional subscale of the voice handicap index (effect sizes = 0.44 and 0.48 for speech and language therapy and surgery groups, respectively) and in the physical symptoms subscale of the voice symptom scale (effect sizes = 0.38 and 0.43 for speech and language therapy and surgery groups, respectively). The largest changes in the individual scales were in the physical aspects of voice subscale of the voice handicap index (effect sizes = 0.71 and 0.81 for speech and language therapy and surgery groups, respectively) and in the voice impairment subscale of the voice symptom scale (effect sizes = 0.78 and 1.00 for speech and language therapy and surgery groups, respectively). The effect sizes corresponding to the change in the total score, respective to speech and language therapy and surgery, for each of the three voice questionnaires, were: vocal performance questionnaire, 1.04 and 0.82; voice handicap index, 0.62 and 0.72; and voice symptom scale, 0.78 and 1.06. The two patient groups were very similar both at baseline and follow up (Figure 1), with no significant differences between them.

The changes in the three voice questionnaires were highly correlated, as follows: vocal performance questionnaire vs voice symptom scale, 0.74 (95 per cent CI: 0.65, 0.81); vocal performance questionnaire vs voice handicap index, 0.76 (95 per cent CI: 0.68, 0.83); and voice symptom scale vs voice handicap index, 0.83 (95 per cent CI: 0.76, 0.87). All differences were significant ($p < 0.0001$). The correlation between the voice symptom scale and the voice handicap index was greater, due in part to the small number of shared items between the two questionnaires. Changes in the subscale components of the voice
The pattern of effect sizes observed in Table I indicates very small changes in the generic health measures. For subjects undergoing speech and language therapy, there were small or medium effects in each grade–roughness–breathiness–asthenia–strain component. For subjects undergoing surgery, the effect sizes were smaller in all components except for roughness, the least sensitive component.

There was little evidence of substantive change in any of the generic health status instruments – all effect sizes were less than 0.3. Thus, although some of the changes were statistically significant, the small effect sizes suggest that they were not clinically important.

**Discussion**

The pattern of effect sizes observed in Table I indicates very small changes in the generic health status. The data suggest that the improvements in quality of life were minimal, with effect sizes typically below 0.3, indicating that the changes were not clinically significant.

**TABLE I**

<table>
<thead>
<tr>
<th>Scale &amp; subscale</th>
<th>Group</th>
<th>Baseline mean (SD)</th>
<th>Follow-up mean (SD)</th>
<th>Mean*</th>
<th>95% CI</th>
<th>Effect size†</th>
<th>t</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>VPQ</td>
<td>SLT</td>
<td>32.3 (9.2)</td>
<td>21.8 (6.8)</td>
<td>10.5</td>
<td>8.3, 12.6</td>
<td>1.04</td>
<td>9.67</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>32.2 (9.2)</td>
<td>22.9 (8.3)</td>
<td>9.3</td>
<td>6.1, 12.4</td>
<td>0.82</td>
<td>5.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>VHI</td>
<td>Physical aspects</td>
<td>21.4 (6.7)</td>
<td>15.3 (8.2)</td>
<td>6.1</td>
<td>4.3, 7.9</td>
<td>0.71</td>
<td>6.72</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>19.4 (7.3)</td>
<td>13.1 (8.2)</td>
<td>6.3</td>
<td>4.2, 8.5</td>
<td>0.81</td>
<td>5.89</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Functional aspects</td>
<td>14.4 (9.0)</td>
<td>9.7 (8.1)</td>
<td>4.8</td>
<td>2.9, 6.6</td>
<td>0.54</td>
<td>5.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>14.6 (9.4)</td>
<td>10.4 (7.4)</td>
<td>4.2</td>
<td>2.3, 6.1</td>
<td>0.61</td>
<td>4.41</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Emotional aspects</td>
<td>11.8 (9.5)</td>
<td>7.9 (8.4)</td>
<td>3.9</td>
<td>2.05, 5.8</td>
<td>0.44</td>
<td>4.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>12.5 (10.4)</td>
<td>7.9 (7.4)</td>
<td>4.5</td>
<td>1.9, 7.1</td>
<td>0.48</td>
<td>3.51</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>SLT</td>
<td>47.8 (22.8)</td>
<td>32.9 (22.7)</td>
<td>14.9</td>
<td>9.9, 19.9</td>
<td>0.62</td>
<td>5.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>46.3 (24.6)</td>
<td>31.4 (21.2)</td>
<td>15.0</td>
<td>9.3, 20.8</td>
<td>0.72</td>
<td>5.24</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>VoiSS</td>
<td>Impairment</td>
<td>30.7 (11.8)</td>
<td>20.1 (11.7)</td>
<td>10.5</td>
<td>7.7, 13.4</td>
<td>0.78</td>
<td>7.27</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>32.2 (11.7)</td>
<td>20.4 (12.3)</td>
<td>11.8</td>
<td>8.5, 15.2</td>
<td>1.00</td>
<td>7.12</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Physical symptoms</td>
<td>9.7 (4.9)</td>
<td>8.1 (5.5)</td>
<td>1.6</td>
<td>0.7, 2.4</td>
<td>0.38</td>
<td>3.52</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>9.0 (5.0)</td>
<td>7.6 (5.2)</td>
<td>1.4</td>
<td>0.5, 2.3</td>
<td>0.43</td>
<td>3.15</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Emotion</td>
<td>8.2 (6.7)</td>
<td>4.7 (5.5)</td>
<td>3.4</td>
<td>2.04, 4.8</td>
<td>0.52</td>
<td>4.92</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>9.3 (8.0)</td>
<td>4.7 (5.5)</td>
<td>4.6</td>
<td>2.8, 6.5</td>
<td>0.69</td>
<td>4.99</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Total</td>
<td>SLT</td>
<td>48.4 (18.1)</td>
<td>32.5 (19.2)</td>
<td>16.0</td>
<td>2.2, 20.4</td>
<td>0.78</td>
<td>7.16</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>50.4 (19.9)</td>
<td>32.4 (19.5)</td>
<td>18.0</td>
<td>13.2, 22.8</td>
<td>1.06</td>
<td>7.57</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>GRBAS</td>
<td>Grade</td>
<td>1.9 (1.1)</td>
<td>1.3 (0.7)</td>
<td>0.7</td>
<td>0.4, 0.9</td>
<td>0.55</td>
<td>5.09</td>
<td>&lt;0.01</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>2.1 (1.1)</td>
<td>1.7 (0.8)</td>
<td>0.4</td>
<td>0.09, 0.8</td>
<td>0.35</td>
<td>2.53</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Roughness</td>
<td>1.1 (0.9)</td>
<td>0.9 (0.8)</td>
<td>0.2</td>
<td>−0.05, 0.4</td>
<td>0.16</td>
<td>1.56</td>
<td>0.12</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>1.2 (0.9)</td>
<td>0.9 (0.8)</td>
<td>0.3</td>
<td>0.01, 0.6</td>
<td>0.29</td>
<td>2.09</td>
<td>0.04</td>
</tr>
<tr>
<td></td>
<td>Breathness</td>
<td>1.5 (0.9)</td>
<td>1.0 (0.6)</td>
<td>0.5</td>
<td>0.3, 0.7</td>
<td>0.57</td>
<td>5.21</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>1.8 (0.9)</td>
<td>1.5 (0.9)</td>
<td>0.3</td>
<td>0.03, 0.6</td>
<td>0.32</td>
<td>2.23</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Asthenia</td>
<td>1.1 (1.0)</td>
<td>0.6 (0.6)</td>
<td>0.5</td>
<td>0.3, 0.7</td>
<td>0.49</td>
<td>4.58</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>1.4 (0.9)</td>
<td>1.2 (0.9)</td>
<td>0.2</td>
<td>−0.03, 0.5</td>
<td>0.25</td>
<td>1.77</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Strain</td>
<td>0.9 (0.9)</td>
<td>0.6 (0.8)</td>
<td>0.3</td>
<td>0.03, 0.6</td>
<td>0.32</td>
<td>2.23</td>
<td>0.03</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>1.1 (0.8)</td>
<td>1.2 (0.9)</td>
<td>0.1</td>
<td>−0.04, 0.2</td>
<td>0.13</td>
<td>0.90</td>
<td>0.37</td>
</tr>
<tr>
<td>SF 36</td>
<td>Physical component</td>
<td>35.2 (14.2)</td>
<td>38.1 (17.1)</td>
<td>2.9</td>
<td>0.6, 5.3</td>
<td>0.26</td>
<td>2.50</td>
<td>0.01</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>41.6 (14.3)</td>
<td>39.5 (15.6)</td>
<td>−2.1</td>
<td>−4.5, 0.4</td>
<td>−0.23</td>
<td>−1.67</td>
<td>0.10</td>
</tr>
<tr>
<td></td>
<td>Mental component</td>
<td>45.7 (12.5)</td>
<td>47.2 (12.1)</td>
<td>1.5</td>
<td>−0.7, 3.7</td>
<td>0.15</td>
<td>1.37</td>
<td>0.18</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>45.7 (12.6)</td>
<td>48.4 (12.0)</td>
<td>2.7</td>
<td>−0.07, 5.4</td>
<td>0.27</td>
<td>1.96</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Depression</td>
<td>4.3 (3.4)</td>
<td>3.7 (3.7)</td>
<td>0.6</td>
<td>−0.02, 1.1</td>
<td>0.21</td>
<td>1.91</td>
<td>0.06</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>4.5 (3.6)</td>
<td>4.1 (3.7)</td>
<td>0.4</td>
<td>−0.5, 1.3</td>
<td>0.12</td>
<td>0.88</td>
<td>0.39</td>
</tr>
<tr>
<td></td>
<td>Anxiety</td>
<td>7.2 (4.3)</td>
<td>6.6 (4.7)</td>
<td>0.6</td>
<td>−0.08, 1.3</td>
<td>0.19</td>
<td>1.77</td>
<td>0.08</td>
</tr>
<tr>
<td></td>
<td>Surg</td>
<td>7.1 (4.0)</td>
<td>6.3 (4.6)</td>
<td>0.8</td>
<td>−0.1, 1.7</td>
<td>0.24</td>
<td>1.71</td>
<td>0.09</td>
</tr>
</tbody>
</table>

Data represent scores for the various measures. *Mean change from baseline in the direction consistent with an improvement in quality of life (QOL) (a negative score implies a deterioration in quality of life). †Effect sizes = improvement in quality of life divided by the standard deviation (SD) of change scores. CI = confidence intervals; VPQ = vocal performance questionnaire; SLT = speech and language therapy; Surg = surgery; VHI = voice handicap index; VoiSS = voice symptom scale; GRBAS = grade–roughness–breathiness–asthenia–strain scale; SF 36 = short form 36; HAD = hospital anxiety and depression scale.
status instruments in comparison with the voice-specific measures. This supports the conclusion of a previous study which assessed voice therapy alone. Following intervention, there were fairly large changes in self-reported, voice-related quality of life. Similar effect sizes were observed across the three self-reported voice questionnaires. The changes in total scores were highly correlated, suggesting that the three self-reported voice questionnaires were detecting changes in the same sets of patients.

When considering the voice handicap index and the voice symptom scale component scores, greatest change was observed in the voice symptom scale impairment subscale and in the voice handicap index physical aspects of voice subscale. Inspection of the individual items that make up these two subscales suggests that they are more or less equivalent; both relate to voice quality. The very high correlation between the changes in these two subscales supports this suggestion. The remaining two voice handicap index components (functional and emotional aspects of health) broadly overlap with the voice symptom scale emotion component; this is reflected in the correlation in change scores between these components. The voice symptom scale physical symptom component has almost no overlap with any of the voice handicap index components; this is reflected in the
much weaker correlations of this subscale with the voice handicap index subscales, and indeed with the other voice symptom scale components.

The grade—roughness—breathiness—asthenia—strain components were less sensitive to change than the self-reported measures. Roughness did not alter much, and was also the only parameter which consistently failed to show any relationship with the total or subscales of any of the three self-reported measures.²

The participants reported comparatively little change in the generic health status questionnaires; a much larger change was observed in the self-reported voice scales. The SF 36 and other generic quality of life instruments have proven validity and reliability, and it is reasonable to suppose that any general, subjective increase in wellbeing would be reflected in these instruments. Similarly, there was little change in the mood variables (the hospital anxiety and depression scales and the SF 36 mental health component score). This suggests that the changes observed in the voice measures were independent of a general improvement in subjective wellbeing as a result of receiving a medical intervention, but rather reflected real improvement in voice-related quality of life. At the same time, the lack of change in the generic measures also highlights the need for a voice-specific questionnaire in assessing voice treatment effectiveness; the SF 36 does not include voice-sensitive components.

### Acknowledgements

This research was supported by a grant from the Wellcome Trust.

### References


Address for correspondence:
Prof. Janet A Wilson,
Dept of Otolaryngology Head Neck Surgery,
Freeman Hospital,
Newcastle upon Tyne NE7 7DN, UK.
Fax: (44) 191 223 1246
E-mail: j.a.wilson@ncl.ac.uk

Professor J A Wilson takes responsibility for the integrity of the content of the paper. Competing interests: None declared